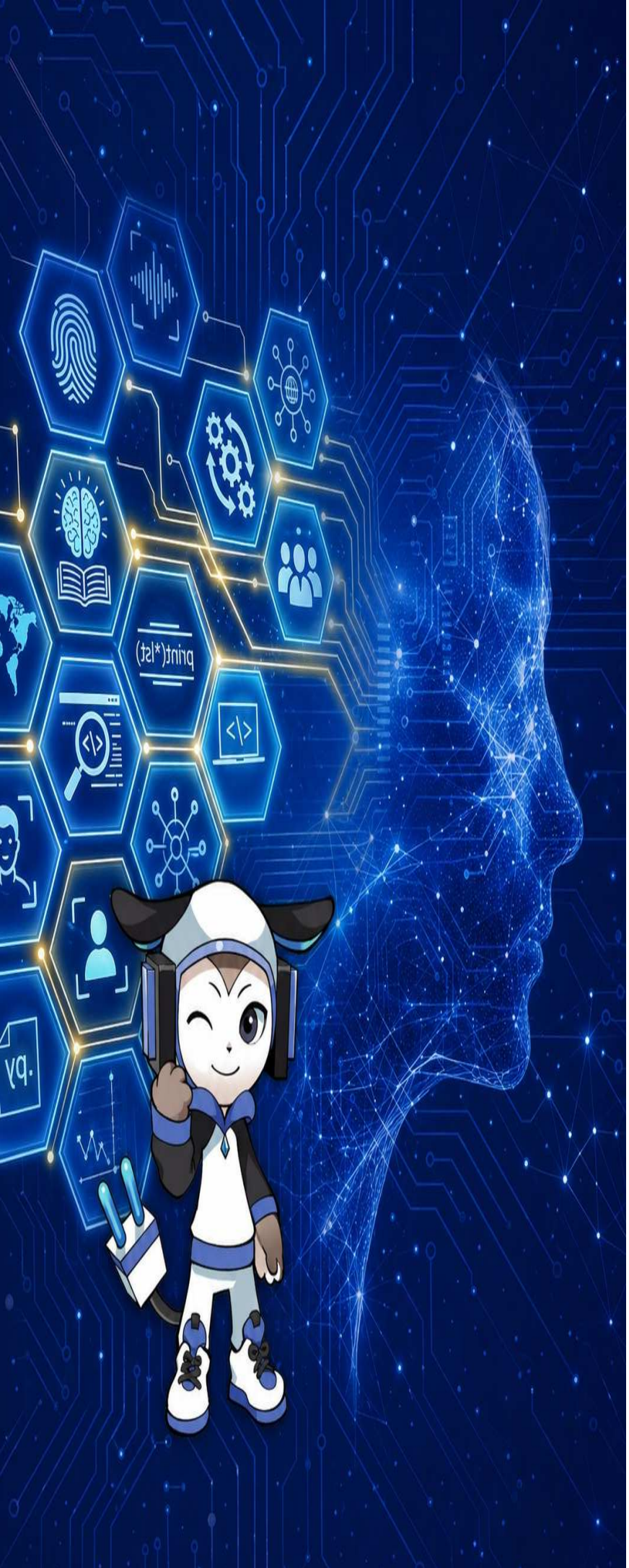


Programming and Artificial Intelligence

An Introduction to
Information and Communication Technology

Egyptian Baccalaureate

2nd Year

Part One

11

SPRIX

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Programming and Artificial Intelligence

Egyptian Baccalaureate

2nd Year

The International Baccalaureate™

has reviewed the pedagogical and assessment approaches underpinning this textbook.

© 2026 Ministry of Education and Technical Education

All rights reserved.

Copyrights and accountability for the content of this textbook remain with the Ministry of Education and Technical Education.

No part of this publication may be reproduced, stored in a retrieval system,

or transmitted in any form or by any means - electronic, mechanical, photocopying, recording,

or otherwise - without the prior written permission of the copyright holders.

Ministry of Education and Technical Education

New Administrative Capital

Cairo, Egypt

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Introduction

In light of Egypt's Vision 2030 and the state's ongoing efforts to develop education, the Ministry of Education and Technical Education places great emphasis on updating school curriculums — particularly in information and communication technology — due to its vital role in fostering critical thinking skills, problem-solving abilities, and a focus on building a deep understanding of fundamental concepts, moving away from rote memorization, especially in the early stages. The content is organized in a gradual and systematic manner that enables students to progress smoothly from simple to complex concepts, linking technical concepts to real-life situations, and relying on interactive methods that include teamwork and discussion, making learning more enjoyable.

In this context, the Ministry has been keen to benefit from successful international experiences. This has led to collaboration between the Egyptian and Japanese sides in developing information and communication technology curriculums, given Japan's global excellence in this field, as evidenced by its notable achievements in international competitions, which reflect the strength of its curriculums and teaching methods.

From this concern, this book is the result of the contribution between the Ministry of Education and Technical Education and the Japanese side.

This book represents a first step within a comprehensive plan to develop information and communication technology curriculums across all grade levels, with the aim of building an integrated educational system that combines the best international practices with Egypt's authentic identity.

The Ministry of Education and Technical Education firmly believes that developing education is a true investment in the nation's future, and that updating curriculums is not an end in itself, but rather a means to build an aware, creative generation, capable of keeping pace with scientific and technological changes in today's world.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Contents

1 **Information Technology and Society** 4

1-1 Development of Information Technology and Social Transformation 4

1-2 How AI Works 12

1-3 AI in Daily Life and Industry 19

1-4 Ethical Issues with AI 26

2 **Cybersecurity** 33

2-1 Cryptographic Technologies and Authentication 33

2-2 Network Security Design 41

2-3 Incident Response and Risk Management 48

3 **Web Applications** 55

3-1 The Overall Structure of Web Applications 55

3-2 Web Application Communication Methods 62

3-3 Fundamentals of Frontend Technology 68

4 **Web and Media Design** 75

4-1 Types and Characteristics of Media 75

4-2 Information Design and User Experience for Websites 81

4-3 Methods for Evaluating Websites 88

4-4 The Iterative Improvement Process for Websites 95

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 1-1

Development of Information Technology and Social Transformation

Learning Objectives

- Explain the major stages in the development of information technology and their impact on society.
- Give examples of social changes and emerging technologies brought about by information technology, and explain their characteristics.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

On an ordinary day, a student in Egypt checks messages on an SNS app, pays for breakfast with a cashless app, joins a lesson through online learning, and orders a book from an e-commerce (EC) shop. Twenty years ago, most of this was not possible. Information technology has developed through a series of stages — from the first computers to cloud computing — and at each stage it has changed the way people communicate, work, learn, and pay.

? Guiding question: How has information technology developed through its major stages, and how has each stage changed society?

✎ Explore (in pairs)

With a partner, look at the five time periods in the table below before you read on. For each period, name one device or service from that era that you still use or hear about today. Then predict which stage changed daily life the most, and give one reason.

Key Fact:
At each stage, information technology introduced a new technology or service and also changed how society communicates, works, and does business.

Key Concepts:

- Moore's Law
- SNS
- e-commerce
- remote work
- online learning
- cashless payment
- edge computing
- autonomous driving
- AR / VR
- quantum computing

📌 Point!

1. The History of Information Technology (IT)

(1) The major stages in the development of information technology are summarized in the table below.

Time period	Major technologies and events	Impact on society
1940s-60s	Birth of the computer (ENIAC, vacuum tubes)	Mainly used for military and scientific computation
1970s-80s	Spread of personal computers (PCs)	Beginning of personal computer use
1990s	Commercialization of the Internet; the Web	Globalization of information; spread of email
2000s	Rise of smartphones (iPhone, etc.)	Explosive spread of mobile Internet
2010s onward	Spread of cloud computing	Large-scale data analysis and AI; "IT as a service" becomes widespread

IT advanced in stages, each smaller and more connected.

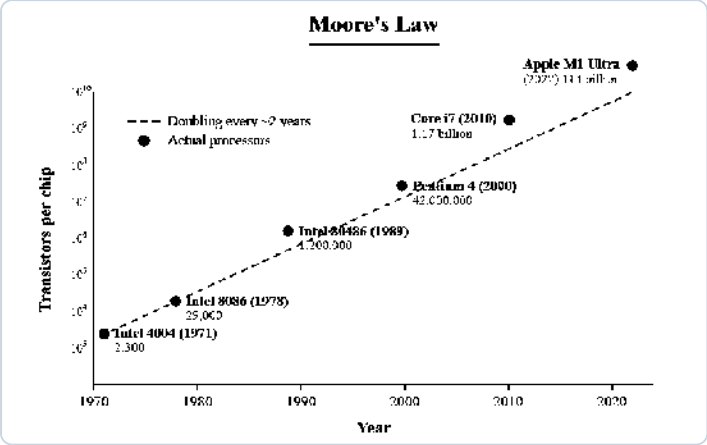
The first computers filled a whole room.

(2) **(Moore's Law)**...the empirical observation that "the number of transistors on an integrated circuit doubles approximately every two years."

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

This remained largely accurate for many years, and computing power increased dramatically.

However, in recent years, miniaturization of transistors has been approaching a physical limit. If circuits are made any smaller, problems arise such as the quantum tunneling effect, in which electrons slip through barriers, and leakage current, in which current escapes unintentionally, making it difficult to achieve both higher performance and lower power consumption at the same time. In response, new directions for performance improvement are being pursued, such as parallel processing using multiple processor cores, and quantum computers based on the principles of quantum mechanics.



Moore's Law: transistor count over time

2. Social Changes Resulting from Information Technology

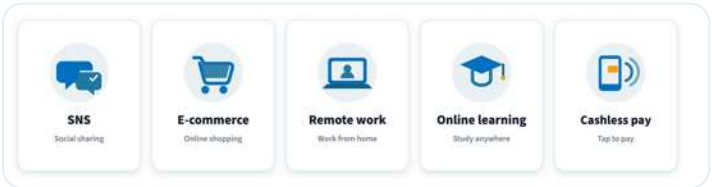
(1) **(SNS (Social Networking Service))**...services that allow users to connect with each other and post and share information. They are highly effective at spreading information rapidly.

(2) **(E-commerce (EC))**...buying and selling goods and services through the Internet.
e.g., online shops such as Amazon and eBay

(3) **(Remote work)**...a working style in which work is performed from home or other remote locations using the Internet.

(4) **(Online learning)**...a learning style in which classes and study materials are delivered using the Internet.

(5) **(Cashless payment)**...a system for making payments using electronic money, QR codes, etc., without using cash.
e.g., credit cards, debit cards, mobile payment apps



Five social changes driven by information technology.

KEY TERMS

Moore's Law — the observation that the number of transistors on a chip roughly doubles about every two years.

This is "the cloud" — real machines in a real place.

PAUSE & THINK

Of these five changes, which would be hardest to give up — and why?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3. Notable Emerging Technologies

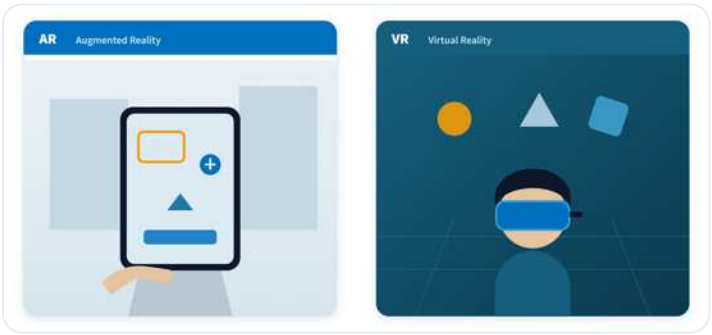
(1) **(Autonomous driving)**...a technology that uses AI to drive a vehicle without human operation. It uses cameras and sensors to recognize the surroundings, makes driving decisions, and controls a vehicle. Because a delay of even 0.1 seconds can lead to an accident, edge computing, in which processing is performed instantly on the vehicle itself rather than sending data to the cloud for judgment, is used.

A self-driving car decides on board (edge computing), without waiting for the cloud.

(2) AR / VR

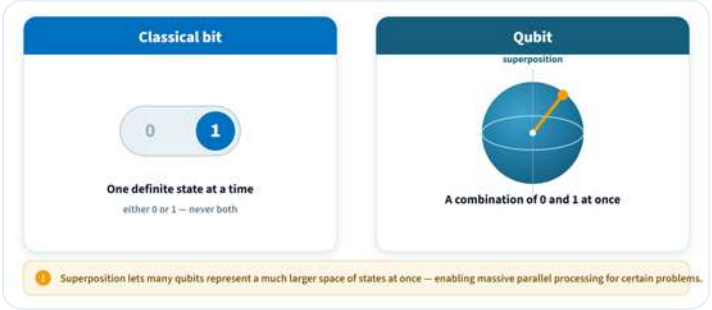
(AR (Augmented Reality))...a technology that overlays digital information on real-world images.

(VR (Virtual Reality))...a technology that allows users to immerse themselves in a virtual space generated by a computer.



AR adds digital layers to the real world; VR replaces it with a virtual one.

(3) **(Quantum computing)**...a technology expected to dramatically speed up computations that are difficult or impossible for traditional computers, by using the principles of quantum mechanics.


Classical bit
0
1
One definite state at a time
either 0 or 1 — never both
Qubit
superposition
A combination of 0 and 1 at once
Superposition lets many qubits represent a much larger space of states at once — enabling massive parallel processing for certain problems.

A classical bit holds one state; a qubit uses superposition.

KEY TERMS

edge computing — processing data on the device itself, instantly, instead of sending it to the cloud.

PAUSE & THINK

In autonomous driving, why is it necessary to process data instantly on the vehicle side using edge computing, rather than sending the data to the cloud for judgment? Explain your answer.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✎ Worked Example

(1) From the following options (A–D), choose the one that lists the stages of information technology (IT) development in the correct chronological order.

A Birth of the computer → Rise of smartphones → Commercialization of the Internet → Spread of cloud computing

B Birth of the computer → Commercialization of the Internet → Rise of smartphones → Spread of cloud computing

C Commercialization of the Internet → Birth of the computer → Spread of cloud computing → Rise of smartphones

D Rise of smartphones → Commercialization of the Internet → Birth of the computer → Spread of cloud computing

(2) Mark each statement (A–D) with ○ if true or × if false.

A Moore's Law is the empirical observation that "the number of transistors on an integrated circuit doubles approximately every two years."

B Moore's Law has been said to be approaching a physical limit in recent years.

C SNS is highly effective at spreading information rapidly.

D E-commerce (EC) refers to purchasing goods at physical stores using cash.

(3) Match each description (a–c) with the most appropriate technology from the choices below (A–C).

a A technology that overlays digital information on real-world images

b A technology that uses AI to drive a vehicle without human operation

c A technology that allows users to immerse themselves in a virtual space generated by a computer

[Choices] **A** Autonomous driving **B** AR **C** VR

✓ Solution

(1) The correct order is "Birth of the computer (1940s-60s) → Commercialization of the Internet (1990s) → Rise of smartphones (2000s) → Spread of cloud computing (2010s onward)." Therefore, **B**.

(2)

A Correct definition of Moore's Law. Therefore, ○.

B It has been said to be approaching a physical limit in recent years. Therefore, ○.

C SNS is a service highly effective at spreading information rapidly. Therefore, ○.

D E-commerce refers to buying and selling through the Internet, not purchases at physical stores using cash. Therefore, ×.

KEY TERMS

Moore's Law — transistors per chip roughly double every two years.

edge computing — data processed on the device itself, instantly.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3)

a: Overlaying digital information on real-world images is AR (Augmented Reality). **B**

b: Driving a vehicle without human operation is autonomous driving. **A**

c: Immersing oneself in a virtual space is VR (Virtual Reality). **C**

Try

1 Answer the following questions.

1. What is the name of the empirical observation that "the number of transistors on an integrated circuit doubles approximately every two years"?
2. What is the term for the buying and selling of goods and services using the Internet?
3. What is the term for the working style in which one works from home or other remote locations using the Internet?
4. What is the term for the system for making payments using electronic money, QR codes, etc., without using cash?
5. What is the term for the technology that uses AI to drive a vehicle without human operation?

2 Answer the following questions.

(1) For each of a–d, indicate which of the following is most closely related: **A** Changes in daily life, **B** Changes in industry and the economy, **C** Changes in healthcare and education.

a Taking classes through online learning.

b Paying for purchases using a mobile payment app on a smartphone.

c Sharing photos with friends on SNS.

d A company introducing a work-from-home system.

(2) From the following options (A–D), choose the one that is NOT an appropriate description of an emerging technology.

A Autonomous driving uses AI to drive a vehicle without human operation.

B AR is a technology that overlays digital information on real-world images.

C VR is a technology that dramatically improves the processing speed of a computer.

D Quantum computing is expected to speed up computations that are difficult for traditional computers.

EXAM-STYLE QUESTION

Analyze how the spread of cloud computing (from the 2010s onward) has changed the way information technology is used. **[6] In your answer, refer to:** large-scale data analysis, AI, and "IT as a service".

KEY TERMS

remote work — working from home or another remote location using the Internet.

online learning — classes and materials delivered over the Internet.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

⚙ Think as an Engineer — Investigate & Decide

Start from the advantage and concern you named in the Pause & Think note, then investigate further and decide.

1 · Collect data. Survey ten classmates: do they usually pay with cash or with a cashless method (card, mobile app, or QR code)? Record the results in a table and identify the most common response.

2 · Analyze the stakeholders. Using your survey and your own reasoning, note one benefit and one drawback of a fully cashless society for each group below.

Group	Benefit	Drawback
A customer paying		
A small shop owner		
A person with no bank card or smartphone		

3 · Decide. Using your survey and your table, decide: should your community move toward cashless payment? Recommend one step and give two reasons.

Stuck? Start with the customer — one benefit is speed of payment; one drawback is needing a card or a phone.

💡 PAUSE & THINK

Cashless payment is spreading in many countries. If a fully cashless society were realized, choose one advantage and one possible concern, and briefly explain the reason for each.

🌍 In a New Context

A village that has never had Internet access is connected to high-speed Internet and cashless payment for the first time. Using the stages and social changes you studied, predict two ways daily life in the village will change, and identify one new problem the village may face.

KEY TERMS

cloud computing — IT delivered as a service over the Internet.

e-commerce — buying and selling through the Internet.

❓ Lesson Question — Answered

The lesson question asked how information technology developed and how each stage changed society. Information technology advanced through a series of stages rather than in a single step, and each stage changed much more than the technology itself. As computing spread, information became global and then mobile, many everyday activities moved online, and information technology grew into a service that supports large-scale data analysis and AI. Each stage also brought social changes in how people communicate, work, learn, and make payments. Overall, the development of information technology is not only about faster machines; at every stage it has reshaped daily life, industry, and the way society handles information.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Computers were invented in the 1940s and were initially used mainly for military and scientific computation. Later, in the 1970s and 1980s, (a) became widespread, allowing individuals to use computers. In the 1990s, (b) was commercialized, advancing the globalization of information. In the 2000s, (c) emerged, leading to the explosive spread of mobile Internet.

1. Fill in the blanks (a)–(c).
2. What technology, which spread from the 2010s onward, supports large-scale data analysis and the use of AI?
3. According to Moore's Law, about how often does the number of transistors on an integrated circuit double?

3 Answer the following questions.

(1) From the following options (A–F), choose all that are emerging technologies.

A Email **B** Autonomous driving **C** Personal computers **D** AR **E** VR **F** Quantum computing

(2) From the following options (A–D), choose the one that most appropriately describes the characteristics of SNS.

A A service that allows users to connect with each other and post and share information, and is highly effective at spreading information rapidly.

B A service for buying and selling goods and services through the Internet.

C A working style in which one works from home or other remote locations.

D A system for making payments using electronic money or QR codes without using cash.

REFLECT & CHALLENGE

Reflect: Which stage of information technology do you think will matter most in the next ten years? Give one reason. Was your prediction at the start of the lesson correct? What changed your mind? **Challenge:** Choose one emerging technology from this lesson (autonomous driving, AR, VR, or quantum computing). Suggest one way it could help solve a real problem, and state one risk.

KEY TERMS

SNS — connects users to post and share information.
cashless payment — payment without cash (e-money, QR).

Key takeaway

Remember:
Information technology developed in stages — computers, the Internet, smartphones, and cloud computing. At each stage it introduced a new technology or service and also changed how society communicates, works, learns, and pays.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Learning Objectives

- Explain what AI is.
- Explain how generative AI is positioned within AI technologies.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

When you use a phone, AI is often at work. A spam filter sorts your email, a store recommends products you might buy, a translation app changes one language into another, and ChatGPT generates text from a prompt. These are all examples of AI, but they are not all the same kind. AI includes machine learning, within machine learning is deep learning, and today's generative AI systems are built on deep learning — a series of nested technologies, each more specialized than the last.

? Guiding question: What is AI, and how are machine learning, deep learning, and generative AI related?

✎ Explore · In Pairs

With a partner, list three tasks a phone or computer does for you that seem to need "intelligence" (for example, sorting spam, recommending a video, or translating text). Then predict: does the computer follow fixed rules, or does it learn from examples? Give one reason.

📌 Point!

1. What is Artificial Intelligence (AI)?

(1) **(AI (Artificial Intelligence))**…a general term for technologies that reproduce or perform intelligent human behavior (learning, reasoning, judgment, etc.) on a computer.
e.g., speech recognition, image recognition, translation

2. The Relationship between AI and Machine Learning

(1) Within AI technologies, there is machine learning; within machine learning, there is deep learning; and within deep learning, there is generative AI. The broad field of AI includes increasingly specialized technologies as a series of nested categories.

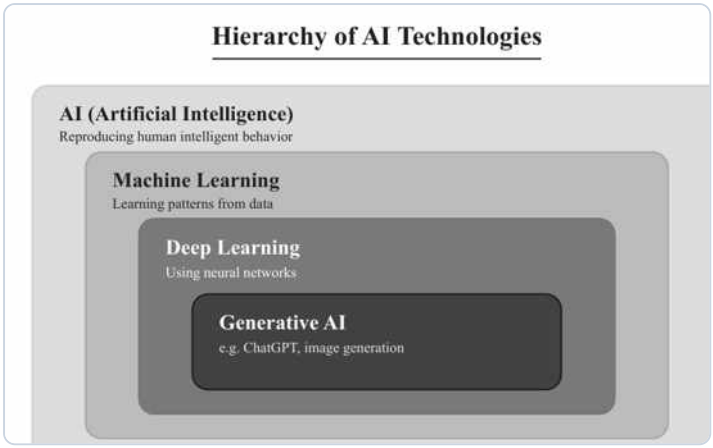
Three everyday AI examples: speech, image, translation.

Today's AI is narrow — expert at one task only.

💡 THINK IT THROUGH

Spam filters and product recommendations are both examples of machine learning, but the tasks they perform are completely different. From the perspective of how AI works, explain what these two applications have in common.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

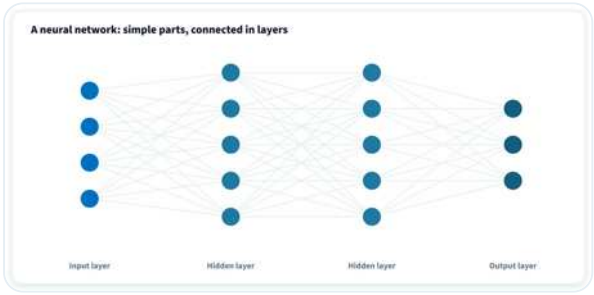


Hierarchy of AI Technologies

(2) **(Machine Learning)**…one of the learning technologies that makes AI work. It learns patterns from data to make predictions and judgments.
e.g., spam filters, product recommendations

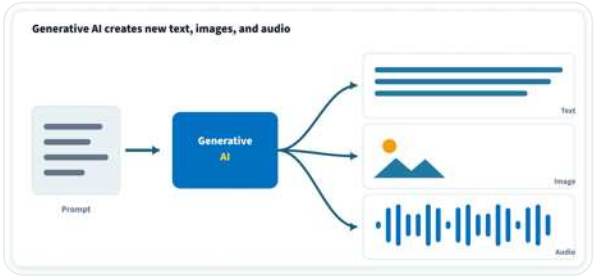
(3) **(Deep Learning)**…an advanced technology within machine learning that uses neural networks. It learns complex patterns using large-scale data.
e.g., image analysis for autonomous driving, speech synthesis

(4) **(Neural network)**…a system modeled after the workings of the nerve cells (neurons) of the human brain. By connecting many components, it learns from data and becomes capable of making complex judgments. It is the core technology supporting recent advances in AI.



A neural network connects many simple components.

(5) **(Generative AI)**…AI technology that uses deep learning to generate new data (text, images, audio, programs, etc.).
e.g., ChatGPT, image generation AIs



Generative AI creates new text, images, and audio.

Fixed hand-written rules versus rules learned from data.

PAUSE & THINK

Deep learning needs large-scale data to learn. Why might an AI struggle with something it has rarely seen in its data?

KEY TERMS

neural network — a system modeled after the nerve cells (neurons) of the human brain, used to learn from data and make complex judgments.

KEY TERMS

generative AI — AI that uses deep learning to generate new data such as text, images, and audio.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Machine learning is one of the technologies that makes AI work.

B Deep learning is a completely different technology from machine learning.

C Generative AI is a technology that uses deep learning to generate new data.

D AI and machine learning have the same meaning.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A technology that learns patterns from data to make predictions and judgments

b A technology that uses neural networks to learn complex patterns

c AI technology that generates new data such as text and images

[Choices] **A** Machine learning **B** Deep learning **C** Generative AI

✓ Solution

(1)

A Machine learning is one of the learning technologies that makes AI work. Therefore, ○.

B Deep learning is a type of machine learning (an advanced technology within machine learning). Therefore, ×.

C Correct description of generative AI. Therefore, ○.

D AI is a broad concept, and machine learning is one of the technologies within it. They do not have the same meaning. Therefore, ×.

(2)

a: Learning patterns from data is machine learning. **A**

b: Using neural networks is deep learning. **B**

c: Generating new data is generative AI. **C**

KEY TERMS

machine learning — learns patterns from data.

deep learning — uses neural networks on large-scale data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the general term, expressed by a two-letter abbreviation, for technologies that reproduce or perform intelligent human behavior on a computer?
2. What is the term for the learning technology that makes AI work by learning patterns from data to make predictions and judgments?
3. What is the term for the advanced technology within machine learning that uses neural networks?
4. What is the term for the AI technology that uses deep learning to generate new data?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that correctly describes the relationship between AI, machine learning, deep learning, and generative AI.

A AI > Deep Learning > Machine Learning > Generative AI

B Machine Learning > AI > Generative AI > Deep Learning

C AI > Machine Learning > Deep Learning > Generative AI

D Generative AI > Deep Learning > Machine Learning > AI

(2) From the following options (A–D), choose the one that is NOT an appropriate example of generative AI.

A ChatGPT

B Image generation AI

C Spam filter

D Audio generation AI

Think as an Engineer — Investigate & Decide

Starting from the Think It Through note about hallucination, investigate further and decide.

1 · Investigate. Find one documented example where a generative AI gave an answer that sounded correct but was actually wrong (a "hallucination"). Record its source — a news article, an official report, or a prompt you tested yourself with a screenshot — and note where you found it. Do not rely on an example recalled from memory without a source.

2 · Weigh it up. Note one benefit and one risk of using generative AI to write a school report — for the student, and for the teacher.

3 · Decide. Should generative AI be allowed for homework in your school? Recommend one rule and give two reasons.

EXAM-STYLE QUESTION

Analyze the relationship between AI, machine learning, deep learning, and generative AI. **[6] In your answer:** use the idea of nested categories, and give one example of each.

THINK IT THROUGH

Generative AI can produce text that sounds plausible but is factually incorrect (a "hallucination"). Explain why it is dangerous to use such output as the answer to a school report as-is, and state what learners should do instead.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? One benefit is speed; one risk is that the answer may be factually incorrect, so it must be checked.

🌍 In a New Context

A farmer wants to use AI to distinguish healthy crops from diseased ones in photographs. Using what you learned about deep learning and image analysis, explain what the AI would need to learn from, and identify one reason its judgment might be wrong.

❓ Lesson Question — Answered

The lesson question asked what AI is and how machine learning, deep learning, and generative AI are related. AI is a general term for technologies that reproduce or perform intelligent human behavior on a computer, and machine learning, deep learning, and generative AI are increasingly specialized fields nested within it rather than parallel to it. The key idea from this lesson is that these are not four separate technologies to learn in isolation, but one broad field understood at different levels of specialization.

KEY TERMS

deep learning — uses neural networks to learn complex patterns.

generative AI — uses deep learning to generate new data.

⭐ REFLECT & CHALLENGE

Reflect: Before this lesson, did you think AI was one single technology? How has your idea changed? Which of the four — AI, machine learning, deep learning, generative AI — was hardest to tell apart, and why?

Challenge: Generative AI can create text, images, and audio. Suggest one helpful use and one possible misuse, and say how the misuse could be reduced.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

The general term for technologies that reproduce or perform intelligent human behavior on a computer is (a). One of the learning technologies that makes (a) work by learning patterns from data is (b). Within (b), an advanced technology that uses neural networks is (c). Furthermore, the technology that uses (c) to generate new data is (d).

1. Fill in the blanks (a)–(d).

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 Answer the following questions.

(1) For each example (a–d), indicate which of the following is most closely related: **A** Machine learning, **B** Deep learning, **C** Generative AI.

a Automatically sorting spam emails

b Recognizing pedestrians on the road in autonomous driving

c Automatically generating new images from text

d Recommending products based on a user's purchase history

(2) From the following options (A–D), choose the one that most appropriately describes the relationship between AI, machine learning, deep learning, and generative AI.

A Machine learning is a type of AI, and deep learning is a type of machine learning.

B AI is a type of machine learning, and generative AI is a type of AI.

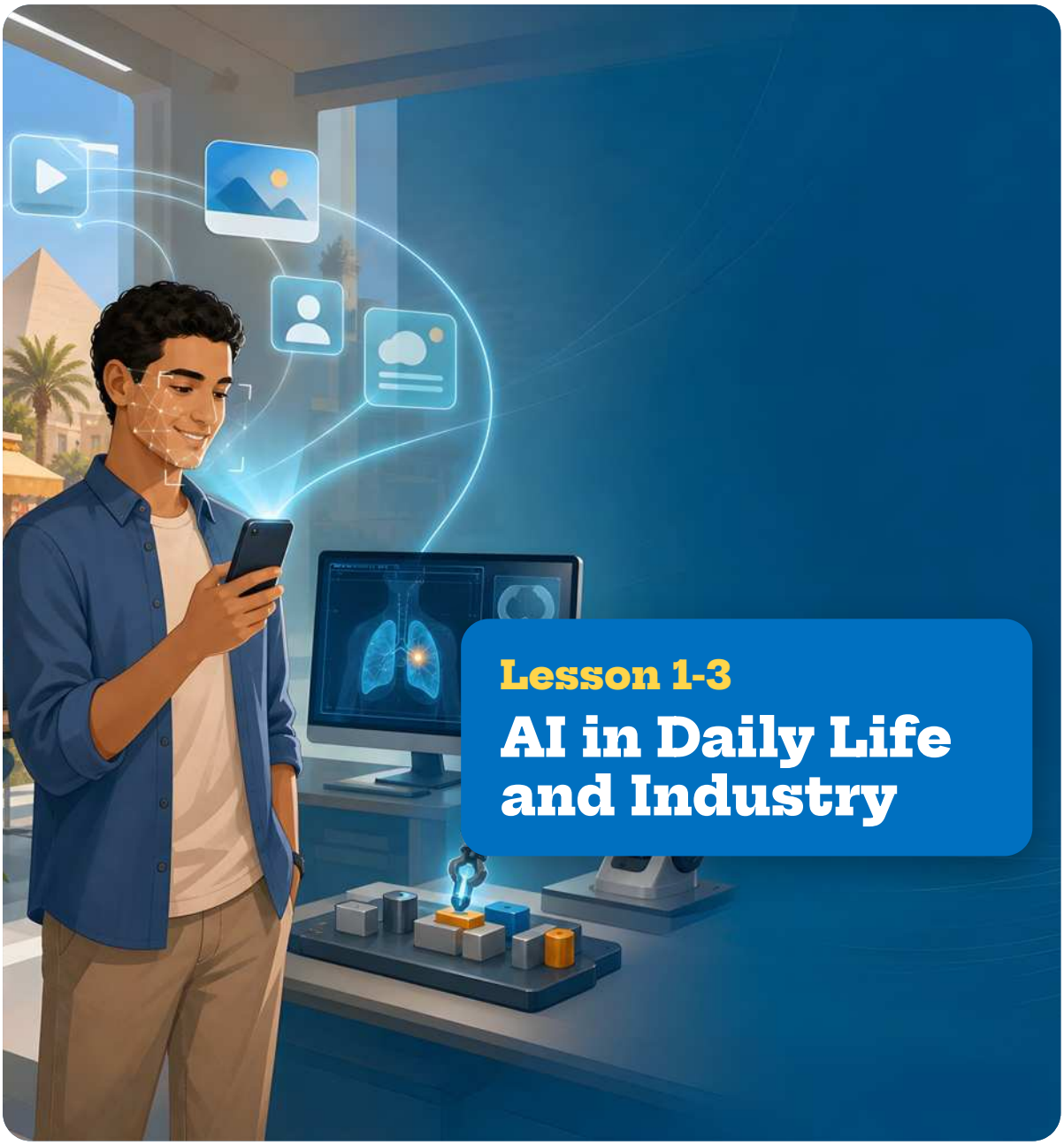
C Deep learning and machine learning are completely different technologies.

D Generative AI is a technology that generates data without using machine learning.

⭐ Key takeaway

Remember:
AI is the broad field; machine learning sits inside it, deep learning inside machine learning, and today's generative AI is built on deep learning — from the broadest idea to the most specialized.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 1-3

AI in Daily Life and Industry

Learning Objectives

- Give examples of how AI is used in daily life and various industries.
- Explain what AI is good at and what requires caution when using AI.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

AI is already part of many everyday services. A video site recommends the next clip, a voice assistant answers a spoken question, a translation tool converts text on a foreign menu, and a phone unlocks by recognizing your face. In industry, AI analyzes medical images, predicts when a machine will fail, and plans delivery routes. AI is effective at identifying patterns in data, but it is not suited to every task, and some uses require caution.

? Guiding question: Where is AI used in daily life and industry, what is AI good at, and what needs caution?

✎ Explore · In Pairs

With a partner, list three services you used this week that you think used AI (for example, a video recommendation, a voice assistant, or a translation). For each, predict what data the AI needed to do its job. Then decide together: which one would matter most if it made a mistake?

Key Fact:
AI is good at finding patterns in data, but human judgment is still needed where ethics, privacy, and responsibility are involved.

Key Concepts:

- recommendation system
- voice assistant
- machine translation
- face recognition
- image-diagnosis AI
- predictive maintenance
- black-box problem
- hallucination

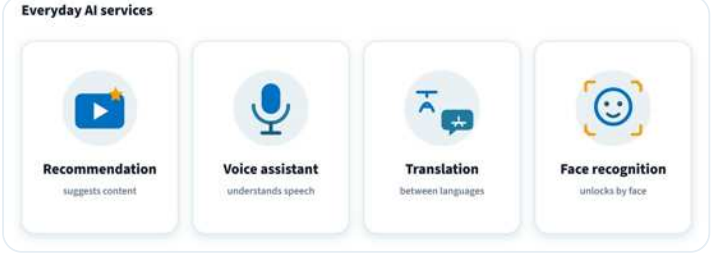
📌 Point!

1. Use of AI in Daily Life

(1) Various AI technologies are already in use around us.

Service	Role of AI	Examples
Recommendation systems	Predict preferences from past behavior data and display recommendations.	YouTube, Amazon, Spotify, etc.
Voice assistants	Recognize a voice, understand commands, and execute them.	Siri, Google Assistant, etc.
Machine translation	Automatically translate text into different languages.	Google Translate, DeepL, etc.
Face recognition	Automatically detect and identify human faces in photographs.	Smartphone unlocking, etc.

Recommenders match new items to your past choices.



Four everyday-life AI services

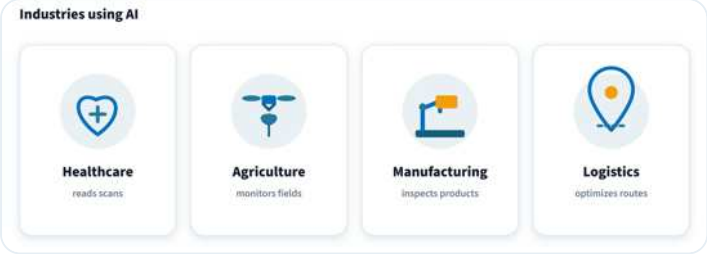
KEY TERMS
recommendation system — a system that predicts preferences from past behavior data and displays recommendations.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

2. Use of AI in Industry

(1) AI technologies are also being introduced in various industries.

Industry	Examples of AI use
Healthcare	Image-diagnosis AI (detecting diseases from X-ray and CT images), drug-discovery support
Agriculture	Predicting harvest timing, detecting pests and diseases
Manufacturing	Automation of product quality inspection, predictive maintenance
Logistics	Optimization of delivery routes



Four industries where AI is used

3. Characteristics of AI and Cautions for Use

(1) What AI is good at

- Finding and classifying patterns in complex data such as images and text
- Recognition and generation of images, audio, and text
- Probabilistic reasoning and prediction based on data

(2) What requires caution when using AI

- Ethical judgments (may lead to discrimination or prejudice)
- Judgments involving personal information or privacy
- Final decision-making (who is responsible for the result)
- Accuracy of judgments when training data is biased
- Generating content that differs from the facts (hallucination)
- Issues of rights when using copyrighted works as training data
- Loss of clarity about how a judgment was made (the black-box problem)

✎ Worked Example

(1) For each of a-d, match the AI technology used in the service with the most appropriate term from the choices below (A–D).

a YouTube recommends videos that match the user's preferences

b Speaking to a smartphone to check the weather

c Translating a foreign-language website into one's native language

d A smartphone camera automatically detects a person's face

[Choices] **A** Recommendation system **B** Voice assistant **C** Machine translation **D** Face recognition

In industry, AI flags what humans might miss.

💡 THINK IT THROUGH

A hospital is considering using image-diagnosis AI to detect diseases from X-ray images. Explain why the final diagnosis should be confirmed by a human doctor rather than left to AI alone.

💡 PAUSE & THINK

AI services often use your personal data — what you watch, buy, or say. Why does this raise privacy concerns, and who should decide how your data is used?

KEY TERMS

black-box problem — when it is unclear how an AI reached its judgment.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) Mark each statement (A–D) with ○ if true or × if false.

A AI is good at probabilistic reasoning and prediction based on data.

B AI is good at recognizing images and audio.

C It is fine to leave ethical judgments entirely to AI.

D AI judgments can become inaccurate when training data is biased.

✓ Solution

(1)

a: Recommendations matched to preferences come from a recommendation system. **A**

b: Operating by speaking is a voice assistant. **B**

c: Translating a foreign language into one's native language is machine translation. **C**

d: Detecting a person's face is face recognition. **D**

(2)

A AI is good at probabilistic reasoning and prediction based on data. Therefore, ○.

B AI is good at recognizing images, audio, and text. Therefore, ○.

C AI judgments can lead to discrimination or prejudice, so ethical judgments should not be left entirely to AI. Therefore, ×.

D Bias in training data affects the accuracy of AI judgments. Therefore, ○.

AI can be confidently wrong — verify its output.

Try

1 Answer the following questions.

1. What is the term for the system that predicts preferences from past behavior data and displays recommendations?
2. What is the term for the AI technology that recognizes a voice, understands commands, and executes them?
3. In which industry are image-diagnosis AIs used?
4. In manufacturing, what is the term for the system that predicts product failures in advance?

EXAM-STYLE QUESTION

Analyze why AI is well suited to some tasks but requires caution in others. **[6] In your answer:** refer to what AI is good at (finding patterns, recognition, prediction) and to what needs caution (ethical judgments, privacy, final decision-making).

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

2 Answer the following questions.

(1) From the following options (A–D), choose the one that does NOT belong to "what requires caution when using AI."

A Ethical judgments

B Finding and classifying patterns

C Judgments involving personal information or privacy

D Final decision-making

(2) For each AI usage example (a–d), indicate which industry is most closely related: **A** Healthcare, **B** Agriculture, **C** Manufacturing, **D** Logistics.

a Optimizing delivery routes

b Detecting diseases from X-ray images

c Detecting pests and diseases

d Automating product quality inspection

⚙ Think as an Engineer — Investigate & Decide

You have just discussed agriculture in the Think It Through note. Now investigate a different industry and decide.

1 · Investigate. Choose one industry from the lesson other than agriculture — healthcare, manufacturing, or logistics. Find one real, documented example of AI being used in it, and record the source you used.

2 · Weigh it up. Note one benefit and one caution of using AI in that industry, using the ideas of what AI is good at and what needs caution.

3 · Decide. Should this industry use AI for this task on its own, or with a human checking the result? Recommend one approach and give two reasons.

Stuck? AI is good at finding patterns quickly; but a human is needed where a wrong result could harm someone or involve responsibility.

🌍 In a New Context

A school library wants to use AI to recommend books to students based on what they have borrowed before. Using what you learned about recommendation systems and what AI is good at, explain how the AI would work, and identify one caution the school should consider.

💡 THINK IT THROUGH

In agriculture, give one realistic advantage of introducing AI and one concern that should be considered before introduction, and briefly explain the reason for each.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

❓ Lesson Question — Answered

AI is used widely in daily life — in recommendation systems, voice assistants, machine translation, and face recognition — and in industry, across fields such as healthcare, agriculture, manufacturing, and logistics. AI is good at finding and classifying patterns in data, recognizing and generating images, audio, and text, and making probabilistic predictions from data. However, it needs caution where ethical judgments, personal information and privacy, and final decision-making are involved, and where training data is biased. So AI is a powerful tool for pattern-based tasks, but human judgment is still needed for responsibility, ethics, and fairness.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Answer the following questions.

(1) Select all options that AI excels at.

A Ethical judgments

B Finding features in images and text and classifying them

C Judgments involving personal information

D Probabilistic reasoning and prediction based on data

E Decision-making that bears responsibility for results

F Recognition and generation of images, audio, and text

(2) From the following options (A–D), choose the one that is NOT an appropriate description of services that use AI.

A A recommendation system predicts preferences from past behavior data and displays recommendations.

B A voice assistant recognizes a voice, understands commands, and executes them.

C Machine translation is a system in which humans manually translate foreign languages.

D Face recognition automatically detects and identifies human faces in photographs.

3 Read the following passage and answer each question.

AI is widely used in our daily lives. For example, recommendations of videos that match one's preferences on a video site come from a (a) system. Also, AI

⭐ REFLECT & CHALLENGE

Reflect: Which use of AI in this lesson surprised you most? Where in your own life do you think AI is making decisions about you? **Challenge:** Choose one task where AI needs caution (ethics, privacy, or final decisions). Suggest one rule that would make using AI for that task safer.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

technology is used in (b), with which one can speak to a smartphone to operate it, and in (c), which converts foreign-language text into one's native language.

1. Fill in the blanks (a)–(c).

(2) From the following (a–d), choose all that require caution when using AI.

a High-speed processing of large amounts of data

b Ethical judgments

c Judgments involving personal information or privacy

d Recognition of images and audio

⭐ Key takeaway

Remember:
AI is good at finding and classifying patterns in data — recommendations, voice, translation, face recognition, and many industry tasks — but human judgment is still needed where ethics, privacy, and responsibility are involved.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 1-4

Ethical Issues with AI

Learning Objectives

- Explain what algorithmic bias is and describe its main causes.
- Connect the basic principles of AI ethics (fairness, transparency, privacy protection, and accountability) to specific situations.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

AI now helps make real decisions about people: selecting candidates for job interviews, identifying individuals through face recognition, and determining how personal data is used. But an AI can be unfair if the data it learned from was biased, and it can be hard to know why it made a decision at all. As AI spreads, questions arise about how to keep it fair, transparent, and accountable, and about who is responsible when it produces an incorrect or unfair result.

? Guiding question: What ethical issues arise as AI spreads, and what principles should guide how we use it?

✎ Explore · In Pairs

With a partner, look at this situation: a face-recognition AI is more likely to misidentify people from some ethnic groups than others. Before you read on, predict who could be harmed by this, and how. Give a reason.

Point!

1. Algorithmic Bias

(1) **(Algorithmic bias)**…bias in AI judgments caused by bias in the training data. *e.g., a hiring AI unfairly evaluating a particular gender; a face-recognition AI being more likely to misidentify particular ethnic groups*

(2) Main causes of bias

- Training data is biased (e.g., insufficient data on certain attributes; past discriminatory tendencies are reflected in the data). Bias can also enter through the way the system is built — for example, an inappropriate choice of variables, a proxy variable that indirectly stands in for a protected attribute, or the design of the model itself.

Key Fact:
If the training data is biased, an AI can reproduce that bias in its judgments; using AI responsibly means checking for bias, being able to explain decisions, and knowing who is accountable.

Key Concepts:

- algorithmic bias
- training data
- face recognition
- privacy
- explainable AI (XAI)
- black-box problem
- responsibility
- fairness
- transparency
- privacy protection
- accountability

KEY TERMS
algorithmic bias — bias in AI judgments caused by bias in the training data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Biased training data leads to biased AI judgments

2. Privacy Issues

(1) The development of AI technology has given rise to new issues concerning **(privacy)**.

Surveillance through face-recognition technology…cameras in public spaces can identify and track individuals.

Mass collection of personal data…large amounts of online behavior data are collected and analyzed.

3. Explainable AI (XAI) and Responsibility

(1) **(Explainable AI (XAI))**…technology that makes it possible for humans to understand why an AI made a particular judgment.
When an AI's decision-making process is opaque (a black box), it is difficult to verify whether the result is correct.

(2) **(Responsibility)**…the question of who is responsible when an AI makes an incorrect judgment.
Views on responsibility differ depending on whether one is the developer, the user, the operator, etc.

4. Basic Principles of AI Ethics

(1) The following basic principles are considered important for using AI appropriately.

Principle	Description
Fairness	Not unjustly discriminating against any particular person or group
Transparency	Showing the AI's decision-making process and inner workings clearly
Privacy protection	Handling personal information appropriately and protecting privacy
Accountability	Being answerable for the AI's decisions

THINK IT THROUGH

When face-recognition technology is used in public spaces, safety and convenience may improve, but privacy concerns also arise. Explain how convenience, safety, and privacy protection should be balanced, and state which basic principle of AI ethics is most closely related to your view.

Privacy: control which personal data is shared.

KEY TERMS

explainable AI (XAI) — technology that makes it possible for humans to understand why an AI made a particular judgment.

Explainable AI reveals the reasons behind a decision.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

The four principles of AI ethics

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Algorithmic bias is bias in AI judgments caused by bias in the training data.

B Even if an AI's decision-making process is opaque, there is no problem as long as the result is correct.

C Explainable AI (XAI) is a technology that makes it possible for humans to understand the AI's reasoning.

D When an AI makes an incorrect judgment, who is responsible has already been clearly determined.

(2) Match each description (a–d) with the most appropriate AI ethics basic principle from the choices below (A–D).

a Not unjustly discriminating against any particular person or group

b Showing the AI's decision-making process and inner workings clearly

c Handling personal information appropriately and protecting privacy

d Being answerable for the AI's decisions

[Choices] **A** Fairness **B** Transparency **C** Privacy protection **D** Accountability

✓ Solution

(1)

A Correct definition of algorithmic bias. Therefore, ○.

B When an AI's decision-making process is opaque, it is difficult to verify whether the result is correct. Therefore, ×.

C Correct description of Explainable AI (XAI). Therefore, ○.

D Views on responsibility when an AI makes an incorrect judgment differ depending on whether one is the developer, the user, the operator, etc., and have not yet been clearly determined. Therefore, ×.

(2)

a: Not discriminating is fairness. **A**

PAUSE & THINK

If no one can explain why an AI rejected someone's job application, is that fair? Which principle of AI ethics is missing?

KEY TERMS

fairness — no unjust discrimination against a person or group.
accountability — being answerable for the AI's decisions.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

b: Showing the decision-making process is transparency. **B**

c: Protecting privacy is privacy protection. **C**

d: Being answerable for decisions is accountability. **D**

Try

1 Answer the following questions.

1. What is the term for the phenomenon in which bias arises in AI judgments due to bias in the training data?
2. What is the term for the technology that makes it possible for humans to understand why an AI made a particular judgment? Also, give its three-letter abbreviation.
3. Among the basic principles of AI ethics, what is the term for the principle that requires not unjustly discriminating against any particular person or group?
4. Among the basic principles of AI ethics, what is the term for the principle that requires handling personal information appropriately?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate cause of algorithmic bias.

A Training data is biased.

B Past discriminatory tendencies are reflected in the data.

C The processing speed of the computer is slow.

D There is insufficient data on certain attributes.

(2) From the following options (A–D), choose all that are privacy issues associated with AI.

A Individuals are identified and tracked through face-recognition technology.

B Online behavior data is collected and analyzed in large quantities.

C AI creates creative works.

D AI computation speed improves.

Think as an Engineer — Investigate & Decide

Building on the hiring-AI case in the Think It Through note, investigate a real example and decide on a rule.

1 · Investigate. Find one documented case where an AI system was accused of being unfair or biased, and record the source you used (a news report, study, or official statement). What group did it affect, and what data might have caused it?

KEY TERMS

privacy — appropriate handling and protection of personal data.

responsibility — who answers when an AI judgment is wrong.

EXAM-STYLE QUESTION

Analyze how algorithmic bias can arise in an AI system, and how it can be reduced. **[6] In your answer:** refer to the cause of bias (biased or insufficient training data) and to how the principles of fairness and transparency help address it.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

2 · Consider responsibility. When an AI makes an unfair decision, who is responsible — the developer, the company using it, or the operator? Note the responsibility each one holds.

3 · Decide. Before a company uses a hiring AI, what one rule should it follow to make the system fairer? Justify your rule using the principles of AI ethics.

Stuck? If the training data reflects past bias, the AI can repeat it — one fix is to check the data and test the AI's results across different groups.

🌍 In a New Context

A school wants to use face-recognition cameras at its gate to record attendance automatically. Identify one practical benefit for the school and one privacy concern. Then, if the camera misidentifies a student, who should be responsible — the school, or the company that made the AI? Give a reason.

❓ Lesson Question — Answered

The lesson question asked what ethical issues arise as AI spreads and what principles should guide how we use it. Because AI now makes decisions that affect people, three concerns stand out: an AI can be unfair when it learns from biased data; it can put people's privacy at risk; and it can be difficult to understand, or to hold anyone answerable for, its decisions. Handling these well takes more than accuracy. The lesson sets out four principles to aim for — fairness, transparency, privacy protection, and accountability — and keeping AI fair, transparent, privacy-respecting, and accountable is what makes its use responsible.

💡 THINK IT THROUGH

A hiring AI was found to rate applicants of a particular gender lower. Using the ideas of algorithmic bias and the basic principles of AI ethics, explain the most likely cause and one possible measure to address it. There can be more than one reasonable view. Give the basis for your answer.

KEY TERMS

transparency — showing the AI's decision-making process clearly.

privacy protection — handling personal information appropriately.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

As AI technology spreads widely in society, attention has been drawn to various ethical issues. For example, there is the issue of (a), in which bias arises in AI judgments due to bias in the training data. Also, for the "black-box" problem—where it is unclear why an AI made a particular judgment—a technology called (

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

b) is being researched. Furthermore, when an AI makes an incorrect judgment, (c) is questioned, but no clear standard has yet been established among developers, users, and operators.

1. Fill in the blanks (a)–(c).

3 Answer the following questions.

(1) For each of a-d, indicate which AI ethics basic principle is most closely related: **A** Fairness, **B** Transparency, **C** Privacy protection, **D** Accountability.

a Ensuring that a hiring AI evaluates fairly regardless of gender.

b Disclosing the AI's decision-making process to users in an easy-to-understand way.

c Not using collected personal data for purposes other than as originally intended.

d Being answerable when an AI's diagnostic result turns out to be wrong.

(2) From the following options (A–D), choose the one that most appropriately describes ethical issues with AI.

A There are no ethical issues with AI.

B Algorithmic bias is caused by the AI's processing speed.

C Bias in training data can cause discriminatory bias in AI judgments.

D Even if an AI's decision-making process is opaque, there is no problem for users.

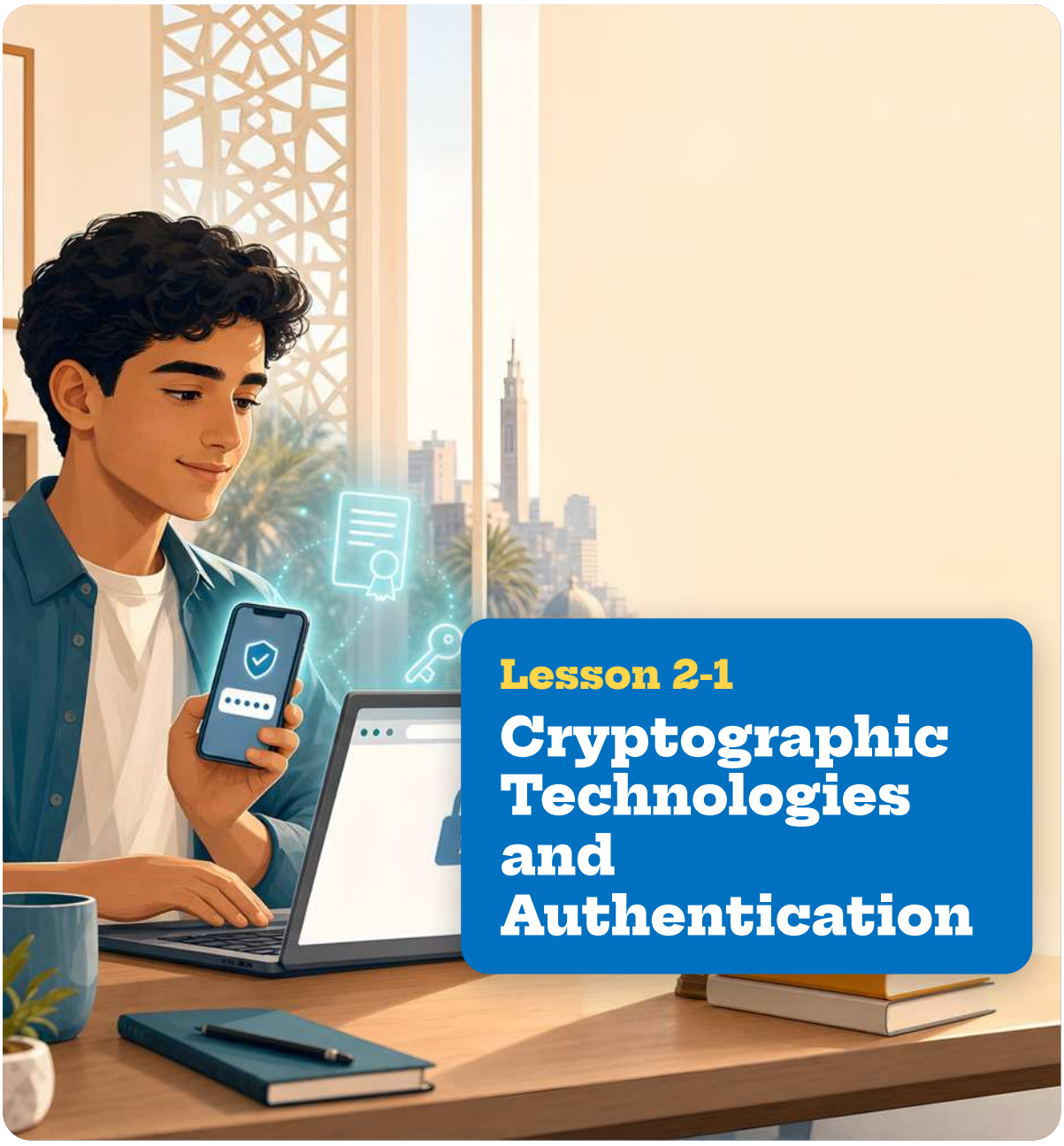
⭐ REFLECT & CHALLENGE

Reflect: Has this lesson changed how much you would trust an AI decision about you? Which of the four principles do you think is hardest to guarantee, and why?
Challenge: Choose one place AI is used (hiring, face recognition, or recommendations). Suggest one rule based on the AI ethics principles that would make it more trustworthy.

⭐ Key takeaway

Remember:
If training data is biased, an AI can reproduce that bias; using AI responsibly means checking for bias, being able to explain decisions, and knowing who is accountable — guided by fairness, transparency, privacy protection, and accountability.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 2-1

Cryptographic Technologies and Authentication

Learning Objectives

- Explain how HTTPS secures communication: how the TLS handshake uses public-key cryptography to share a key safely, and how common-key cryptography then protects the data that follows.
- Describe how authentication, including multi-factor authentication, is used in real services, and explain why security technologies are combined.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Every time you log in to a bank app, buy something online, or open a website with a small padlock in the address bar, hidden technologies are working to keep your information safe. Messages are scrambled so no one else can read them, the website proves that it is genuine, and your identity is checked before you are granted access. Without these protections, passwords could be stolen and private data could be read by unauthorized people.

? Guiding question: How do cryptographic technologies and authentication work together to keep online communication and services secure?

Point!

1. How HTTPS Communication Works

(1) **(HTTPS (HTTP Secure))**…a communication protocol that adds TLS encryption to HTTP. By providing encryption, detection of tampering, and verification of the communication partner, it protects users from threats such as eavesdropping, tampering, and impersonation. The procedure for establishing a secure communication is called the TLS handshake.

(2) HTTPS combines common-key (symmetric-key) cryptography and public-key cryptography as follows, taking advantage of the strengths of both methods.

Stage	What happens	Cryptographic method used
① **Connection**	The web server sends a digital certificate and a public key to the browser. The browser verifies the server's authenticity through a certificate authority (CA).	Public-key cryptography
② **Key exchange**	The browser creates a common key for communication, encrypts it with the server's public key, and sends it. The server decrypts it with its private key to obtain the common key.	Public-key cryptography
③ **Data communication**	Data is encrypted and exchanged using the shared common key.	Common-key cryptography

Note: the exchange above shows one way a common key can be shared. Current versions of TLS use Diffie-Hellman key agreement instead.

Key Fact:
Secure online communication and services depend on combining cryptography and authentication — not on any single technology.

Key Concepts:

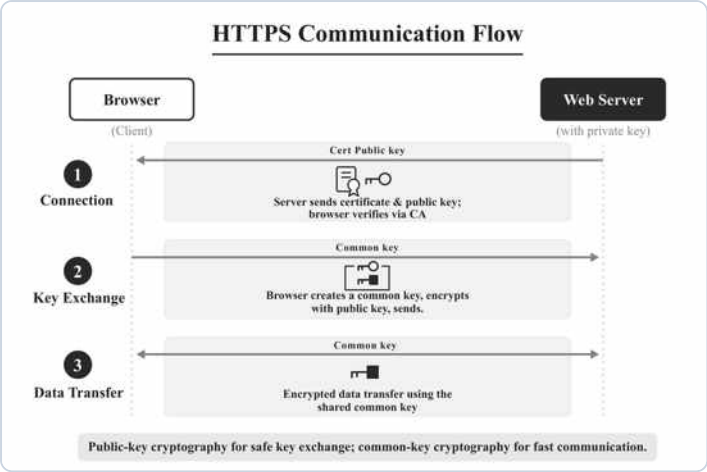
- HTTPS
- common-key cryptography
- public-key cryptography
- TLS handshake
- multi-factor authentication
- digital signature
- digital certificate
- non-repudiation

KEY TERMS

HTTPS — HTTP with TLS encryption added (encryption, tamper detection, and partner verification).

A public key locks; a different private key unlocks. In a digital signature the direction reverses: the private key signs and the public key verifies.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



HTTPS Communication Flow

(3) Public-key cryptography is used to "share keys safely," and common-key cryptography is used to "exchange large amounts of data quickly." This division of roles makes it possible to achieve both security and communication speed.

2. Use of Authentication in Real Services

(1) Many services adopt **(multi-factor authentication)**, in which two or more of the three factors of authentication (knowledge, possession, biometric) are combined.

(2) When only one factor is used, an unauthorized party can gain access if that factor is broken. By combining multiple factors, even if one is broken, another can still prevent access.
e.g., If only a password (knowledge) is used, an unauthorized login is possible once the password is leaked. However, if a notification to a smartphone (possession) is also required, an attacker cannot log in without the device, even if the password is leaked.

(3) The following examples of multi-factor authentication are used in various services.

Service	Examples of authentication used
Online banking	Password (knowledge) + one-time password (possession)
SNS login	Password (knowledge) + SMS authentication (possession)
Company laptop login	Smart card (possession) + fingerprint authentication (biometric)
Online shopping	Password (knowledge) + one-time code from 3-D Secure (possession)

3. Combinations of Security Technologies

(1) Safe communication and services are achieved not by a single technology, but by combining multiple security technologies.

Strong login combines two or more different factors.

KEY TERMS
multi-factor authentication — combining two or more of the three factors: knowledge, possession, biometric.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Technology	Effect
Encryption (common-key, public-key)	Prevents communication content from being read by third parties
Digital signature	Detects data tampering or impersonation of the sender, and makes it impossible for the sender to deny having sent the data afterward (non-repudiation)
Digital certificate	Verifies that the communication partner is authentic
Multi-factor authentication	Prevents unauthorized logins

(2) For example, when purchasing a product through online shopping, these technologies are combined and used as follows.

When you access the site, HTTPS communication begins, and **(encryption)** prevents communication content from being read by third parties.

The browser checks the **(digital certificate)** to confirm that the destination is the genuine shopping site.

For account login, **(multi-factor authentication)** combining a password and SMS authentication, etc., prevents unauthorized logins.

When sending order details, a **(digital signature)** can also be used so that tampering during communication can be detected.

(3) By combining multiple technologies in this way, various threats such as "eavesdropping," "tampering," "impersonation," and "unauthorized login" can be addressed simultaneously.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A HTTPS communication uses both public-key cryptography and common-key cryptography.

B HTTPS communication exchanges all data using public-key cryptography.

C Multi-factor authentication is a method that combines two or more of the three factors of authentication.

D An example of multi-factor authentication is setting two passwords.

(2) Match each threat (a–c) with the most appropriate security technology from the choices below (A–C).

a I want to prevent communication content from being read by third parties.

b I want to verify that data has not been altered during communication.

KEY TERMS

digital signature — detects tampering and impersonation, and provides non-repudiation.

A certificate authority vouches that a site is genuine.

💡 PAUSE & THINK

If a password is leaked, why can a second factor sent to a phone still keep the account safe?

KEY TERMS

public-key cryptography — shares a key safely.
common-key cryptography — exchanges data quickly.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

c I want to verify that the communication partner is authentic.

[Choices] **A** Encryption **B** Digital signature **C** Digital certificate

✓ Solution

(1)

A HTTPS communication uses public-key cryptography for key exchange and common-key cryptography for data communication. Therefore, ○.

B Exchanging all data using public-key cryptography would be slow, so common-key cryptography is used for data communication. Therefore, ×.

C Correct definition of multi-factor authentication. Therefore, ○.

D Even if two passwords are set, both are "knowledge" factors, so this does not constitute multi-factor authentication (it is two-step authentication using the same factor). Therefore, ×.

(2)

a: Preventing reading of communication content is encryption. **A**

b: Detecting data tampering is a digital signature. **B**

c: Verifying the authenticity of the communication partner is a digital certificate. **C**

Try

1 Answer the following questions.

1. In HTTPS communication, what cryptographic method is used so that the browser and server can safely share a common key?
2. In HTTPS communication, what cryptographic method is used for data communication after the common key has been shared?
3. Briefly explain the reason for combining the cryptographic methods in (1) and (2), based on the characteristics of each method.

EXAM-STYLE QUESTION

Explain why HTTPS uses public-key cryptography to share the key but common-key cryptography to exchange the data. **[6] Refer to:** security and speed.

2 Answer the following questions.

(1) For each service (a–d), indicate which combination of authentication factors is used: **A** Knowledge + possession, **B** Knowledge + biometric, **C** Biometric + possession.

a Password + one-time password

b Password + fingerprint authentication

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

c Password + SMS authentication

d Face recognition + smart card

(2) From the following options (A–D), choose the one that is NOT an appropriate description of combinations of security technologies.

A Encryption is a technology that prevents communication content from being read by third parties.

B Digital signatures are a technology that improves communication speed.

C Digital certificates are a system for verifying that the communication partner is authentic.

D By combining multiple security technologies, various threats can be addressed.

Think as an Engineer — Investigate & Decide

A school in your city is building an online portal where students check their grades and pay fees. You must choose how students log in. Which is the better choice for this service?

☐ **A password only** ☐ **A password plus a one-time password sent to the phone**

1 · Collect data. Ask ten classmates which login methods they use and trust (password, one-time password, SMS, fingerprint or face). Record the results in a table and find the most common answer.

2 · Analyze the stakeholders. For each group — a student, the school, and a student with no smartphone — note one benefit and one drawback of requiring a second factor.

3 · Decide. Recommend one login method for the portal and give two reasons based on your survey and your table. **In pairs,** compare your recommendations and agree together on the single strongest reason before you write it down.

Give evidence for your answer.

Stuck? Start with the user: a password alone fails once it leaks; adding a possession factor (a one-time password) still blocks the login.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

🌍 In a New Context

A small online shop wants to secure both customer login and payment. Using what you learned, name two security technologies it should use and say where each one helps, then identify one threat that could still remain.

❓ Lesson Question — Answered

The lesson asked how cryptographic technologies and authentication keep communication and services secure. HTTPS combines public-key cryptography — used to share a key safely — with common-key cryptography — used to exchange data quickly — so communication is both secure and fast. Beyond communication, services stay safe by combining several technologies: encryption hides content, digital certificates prove a site is genuine, digital signatures detect tampering, and multi-factor authentication blocks unauthorized logins. Security online therefore comes from layering these technologies so that many threats are addressed at the same time.

KEY TERMS
digital certificate — proves a site is genuine.
multi-factor authentication — blocks unauthorized logins.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

HTTPS communication, used to access websites safely, combines multiple cryptographic technologies. First, the web server sends a (a) and a public key to the browser. The browser verifies whether the server is authentic through (b). Next, the browser creates a (c) for communication and encrypts it with the server's public key before sending. After this, all data communication is encrypted using this (c).

1. Fill in the blanks (a)–(c).
2. Briefly explain the reason for using common-key cryptography rather than public-key cryptography for data communication.

⭐ REFLECT & CHALLENGE

Reflect: Which of these protections do you rely on most when you go online?
Challenge: Suggest one extra security measure an online shop could add, and one threat it would still not remove.

KEY TERMS
encryption — keeps content private.
digital signature — detects tampering; gives non-repudiation.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 Answer the following questions.

(1) For each threat (a–d), indicate which technology is primarily used to address it: **A** Encryption, **B** Digital signature, **C** Digital certificate, **D** Multi-factor authentication.

a A password is leaked and unauthorized login occurs.

b Data is altered during communication.

c A user is led to a fake website.

d Communication content is intercepted by a third party.

(2) From the following options (A–D), choose the one that is NOT an appropriate example of multi-factor authentication.

A Password + smartphone notification

B Password + secret question

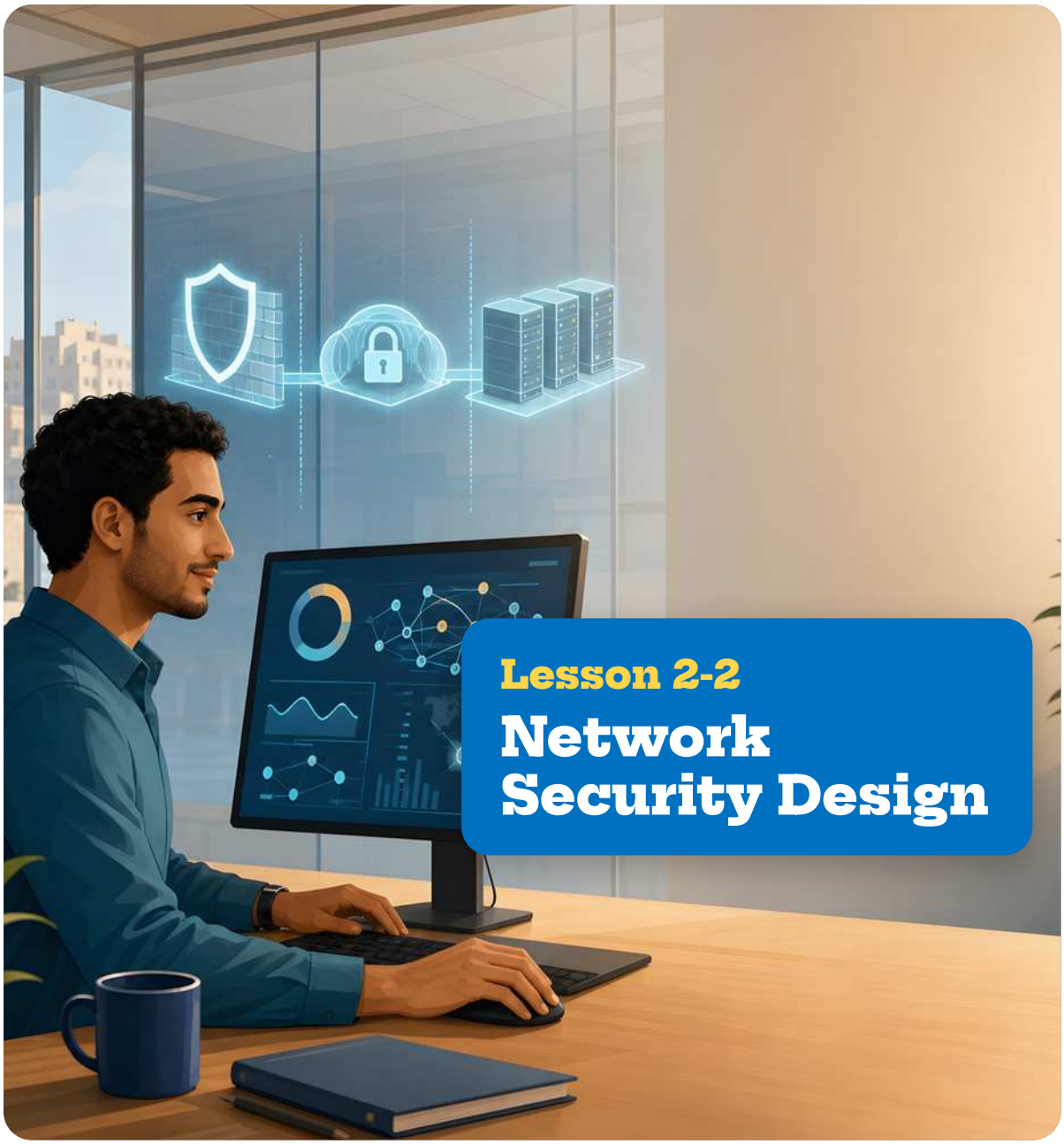
C Fingerprint authentication + smart card

D Password + SMS authentication

⭐ Key takeaway

Remember:
HTTPS uses public-key cryptography to share a key safely and common-key cryptography to move data quickly. Services stay safe by layering encryption, certificates, signatures, and multi-factor authentication.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 2-2

Network Security Design

Learning Objectives

- Explain the roles of firewalls, VPNs, and DMZs in protecting a network.
- Describe defense in depth and zero trust, and explain why layered, verified access is stronger than a single barrier.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

An organization's network holds confidential data — customer records, financial files, and private messages. Attackers constantly look for a way in. A single lock on the front door is not enough: if it is picked, everything inside is exposed. For this reason, modern network security uses several layers, so that even when one defense is breached, the next one still protects the system.

? Guiding question: How can an organization design its network so that, even if one defense is breached, the system stays protected?

Point!

1. The Role of Firewalls

(1) **(Firewall)**…a system installed at the entry/exit point of a network that allows or denies communication based on rules set in advance.
e.g., Allow access to the web server, but deny direct external access to the database server.

2. VPN (Virtual Private Network)

(1) **(VPN)**…a technology that creates an encrypted virtual private line (tunnel) over the Internet, allowing safe connection to a network from a remote location.

(2) Main uses of VPN

Remote work…connecting safely to an organization's network from home or while away.

Connecting branches…connecting offices in different locations safely.

(3) Because communication content is encrypted when using a VPN, the risk of eavesdropping is reduced even when using public Wi-Fi.

3. DMZ (Demilitarized Zone)

(1) **(DMZ)**…an area where servers exposed to the external network (web servers, mail servers, etc.) are placed separately from the internal network.

(2) By providing a DMZ, even if a public-facing server is attacked, damage to the internal network can be prevented.

Key Fact:
Strong network security comes from layering measures — firewall, VPN, DMZ, endpoint protection — and from verifying every access (zero trust), not from a single barrier.

Key Concepts:

- firewall
- VPN
- DMZ
- defense in depth
- zero trust
- perimeter

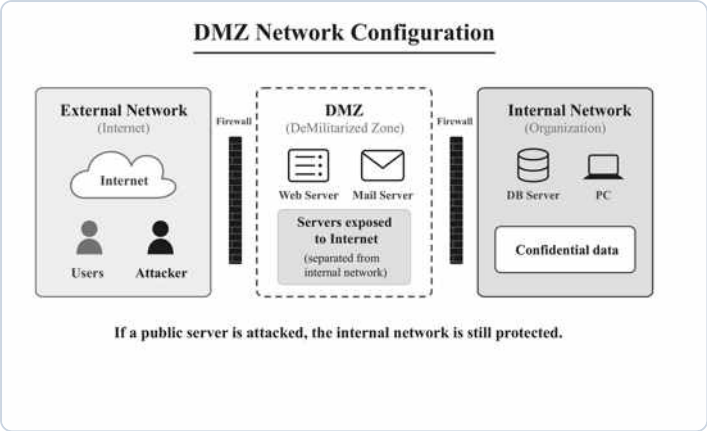
KEY TERMS
firewall — allows or denies communication at a network's entry/exit point based on preset rules.

A firewall allows or blocks traffic by rules.

A VPN tunnels data safely through a hostile network.

KEY TERMS
DMZ — a zone for public-facing servers, separated from the internal network.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



DMZ Network Configuration

4. Defense in Depth

(1) **(Defense in depth)**…the concept of layering multiple security measures so that, even if one measure is breached, the next can still protect the system.

(2) In network security, multiple measures are combined as follows.

Layer	Examples of measures
Layer 1: Network entry	Block unauthorized communication using a firewall
Layer 2: Communication path	Encrypt communication using a VPN
Layer 3: Server placement	Separate public-facing servers from the internal network using a DMZ
Layer 4: Endpoints	Install antivirus software and keep the OS updated

5. The Concept of Zero Trust

(1) **(Zero trust)**…a security concept that means the system does not trust any communication, including from within the organization's network, and always verifies it.

e.g., Even for access from inside the organization, identity verification and access permission checks are performed every time.

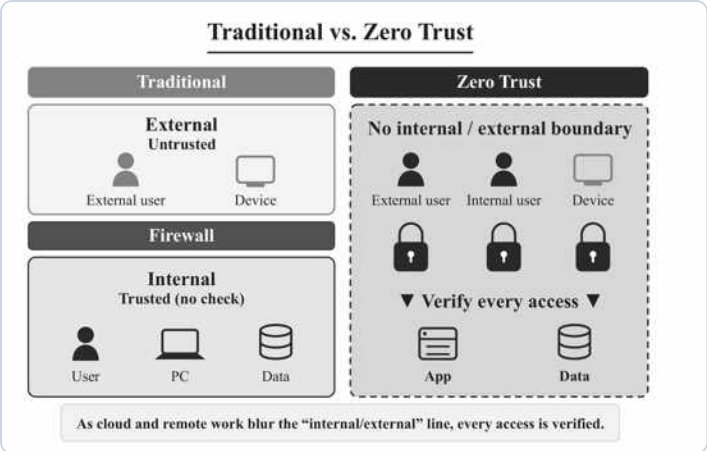
(2) Conventional network security was designed mainly around protecting a perimeter, based on the premise that "inside the organization's network is safe, and outside is dangerous." However, with the spread of cloud services and remote work, the boundary between "inside" and "outside" has become blurred, and designs that protect only the perimeter have become insufficient.

PAUSE & THINK

Why is it safer to place a public web server in a DMZ rather than inside the internal network?

Defense in depth: if one layer falls, the next holds.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Traditional vs. Zero Trust

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A firewall allows or denies communication based on rules set in advance.

B Using a VPN can reduce the risk of eavesdropping even on public Wi-Fi.

C A DMZ is a system that exposes all servers to the outside.

D Defense in depth is the concept of layering multiple security measures.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A technology that creates an encrypted virtual private line over the Internet for safe connection from a remote location

b An area where servers exposed to the outside are placed separately from the internal network

c A system that allows or denies communication based on rules set in advance

[Choices] **A** Firewall **B** VPN **C** DMZ

✓ Solution

(1)

A A firewall controls communication based on rules. Therefore, ○.

B A VPN encrypts communication, reducing the risk of eavesdropping on public Wi-Fi. Therefore, ○.

C A DMZ is an area where public-facing servers are placed separately from the internal network, not a system that exposes all servers. Therefore, ×.

KEY TERMS

defense in depth — layering measures.

zero trust — verify every access.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

D Correct definition of defense in depth. Therefore, ○.

(2)

a: An encrypted virtual private line is a VPN. **B**

b: An area where public-facing servers are isolated is a DMZ. **C**

c: A system that allows or denies communication is a firewall. **A**

Try

1 Answer the following questions.

1. What is the term for the system installed at the entry/exit point of a network that allows or denies communication based on rules?
2. What is the three-letter abbreviation for the technology that creates an encrypted virtual private line over the Internet for safe communication?
3. What is the three-letter abbreviation for the area where servers exposed to the outside are placed separately from the internal network?
4. What is the term for the concept of layering multiple security measures?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate use of a VPN.

A Connecting safely from home to an organization's network

B Connecting offices in different locations safely

C Automatically deleting files infected with viruses

D Encrypting communication content on public Wi-Fi

(2) For each of the security measures (a–d), indicate which layer of defense in depth (A–D) it belongs to.

a Blocking unauthorized communication with a firewall

b Encrypting communication with a VPN

c Installing antivirus software

d Separating public-facing servers from the internal network using a DMZ

[Layers] **A** Network entry **B** Communication path **C** Server placement **D** Endpoints

Think as an Engineer — Investigate & Decide

A company in your city is setting up its office network. It has a public website, staff who work from home, and confidential customer data. You advise on the security design.

EXAM-STYLE QUESTION

Explain why layering several security measures (defense in depth) protects an organization better than relying on a firewall alone. **[6] Refer to:** what happens if one measure is breached.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

☐ A firewall only

☐ A layered design (firewall + VPN + DMZ + endpoints) that also verifies every access (zero trust)

1 · Collect data. List the company's three assets (public website, remote staff, customer database) and the main threat to each.

2 · Rank the layers. For this company, rank the four defense-in-depth layers (network entry, communication path, server placement, endpoints) from most to least critical, and justify your top choice.

3 · Decide. Recommend a design and give two reasons why layering is safer than a single firewall. **In pairs,** compare your designs and agree together on the strongest reason before you write it down.

Give evidence for your answer.

Stuck? Start with the most exposed asset — the public website — and place it in a DMZ so that an attack there cannot reach the internal database.

🌍 In a New Context

A hospital moves its patient records to a cloud service and lets doctors access them from home. Using zero trust and defense in depth, suggest two measures it should apply and say where each one helps, then identify one risk that could still remain.

❓ Lesson Question — Answered

The lesson asked how to design a network so that one breach does not expose everything. The answer is layering. A firewall controls what enters and leaves the network; a VPN encrypts communication from remote locations; a DMZ separates public-facing servers so that an attack on them cannot reach internal data; and endpoint measures protect each device. Because cloud services and remote work have blurred the line between "inside" and "outside," zero trust adds a further principle: no access is trusted automatically, and every request is verified. Together, defense in depth and zero trust keep the network protected even when a single measure fails.

KEY TERMS

perimeter — the old "inside is safe" boundary that cloud and remote work have blurred.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

To protect an organization's network, it is important to combine multiple security measures. First, a (a) is installed at the entry/exit point of the network to block unauthorized communication based on rules. Web servers and other servers exposed to the outside are placed in a (b), separated from the internal network. To connect safely from outside, such as with remote work, communication is encrypted using a (c). The concept of layering multiple measures in this way is called (d).

1. Fill in the blanks (a)–(d).

3 Answer the following questions.

(1) From the following options (A–D), choose the most appropriate reason for providing a DMZ.

A To improve communication speed.

B To expose all servers to the outside.

C To prevent damage to the internal network even if a public-facing server is attacked.

D To automatically detect and remove viruses.

(2) From the following options (A–D), choose the one that most appropriately describes the concept of zero trust.

A Access from within the organization's network is safe, so verification is unnecessary.

B A firewall alone is sufficient to protect the network.

C No communication is trusted and is always verified.

D Only access from outside needs to be monitored.

⭐ REFLECT & CHALLENGE

Reflect: Which layer of defense do you think is easiest for an attacker to get past, and why?
Challenge: Suggest one measure a small company could add beyond a firewall, and one threat it would still not remove.

⭐ Key takeaway

Remember:
Layer your defenses — firewall, VPN, DMZ, endpoints — and verify every access (zero trust). One breach is then far less likely to expose the whole network.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 2-3

Incident Response and Risk Management

Learning Objectives

- Explain what a security incident is and describe the stages of incident response in the correct order.
- Apply the idea of risk assessment by prioritizing risks according to their impact and likelihood.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Even a well-protected system can be attacked. When a security incident happens — data leaked, files tampered with, or a service knocked offline — a rushed and unplanned reaction only makes the damage worse. Organizations follow a clear, ordered response so that damage is contained, the cause is removed, and the same problem does not return. They also decide in advance which risks to guard against first, because no team can fix everything at once.

? Guiding question: When a security incident happens, how should an organization respond in order — and how does it decide which risks to address first?

Key Fact:
Incident response follows six ordered stages, and risk assessment prioritizes risks by impact × likelihood — so an organization acts in the right order and on the right risks first.

Key Concepts:

- security incident
- the six stages
- containment / eradication
- risk = impact × likelihood

KEY TERMS

security incident — a situation in which actual damage occurs due to an information security problem.

Point!

1. What is a Security Incident

(1) **(Security incident)**…a situation in which actual damage occurs due to an information security problem.

(2) Typical examples of security incidents

Type	Description
Information leak	Personal information or confidential information is leaked to the outside
Data tampering	Data within a system is altered without authorization
Service outage	A service becomes unavailable due to attacks on the server, etc.

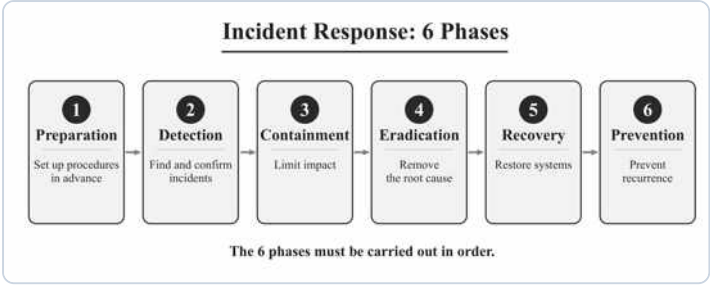
2. The Flow of Incident Response

(1) When a security incident occurs, response proceeds through the following six stages.

Stage	Description
① **Preparation**	Establish response procedures and structures in advance, in preparation for incidents
② **Detection**	Discover and confirm that an incident has occurred
③ **Containment**	Limit the scope of impact so that damage does not spread further
④ **Eradication**	Remove the cause of the incident (e.g., remove malware)
⑤ **Recovery**	Return systems and services to their normal state
⑥ **Prevention of recurrence**	Analyze the cause and take measures so that the same incident does not occur again

Contain first: isolate the infected machine.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Incident Response: 6 Phases

(2) It is important to proceed through the six stages of incident response in order. For example, if recovery is performed without identifying the cause, the same incident may occur again.

3. The Concept of Risk Assessment

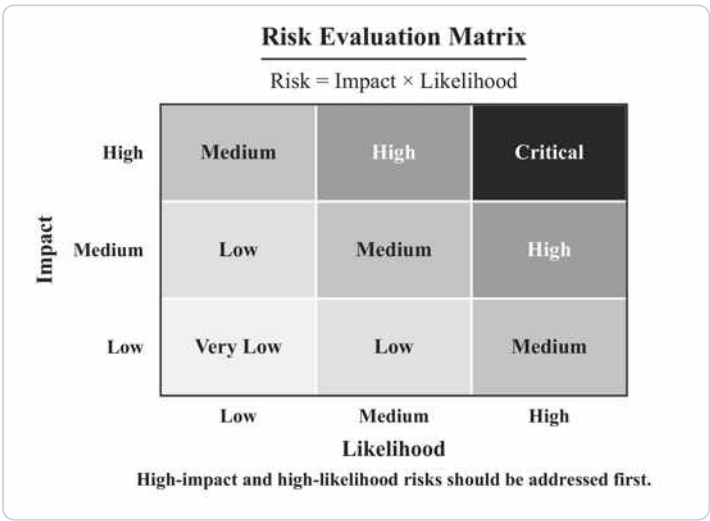
(1) **(Risk)**…the possibility of an information security problem occurring in the future, and its impact.

(2) The size of a risk is determined by the following two factors.

$$\text{Risk size} = \text{Impact} \times \text{Likelihood}$$

(Impact)…how much damage occurs if the risk actually materializes.

(Likelihood)…how high the probability is that the risk will actually materialize.



Risk Evaluation Matrix

(3) Risks with both high impact and high likelihood should be addressed first. Since it is difficult to address all risks at the same time, it is important to set priorities and take measures systematically.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A security incident is a situation in which actual damage occurs due to an information security problem.

Recover: restore a clean system from backup.

PAUSE & THINK

Why must containment come before eradication in the six-stage response?

Rank risks by impact and likelihood; fix the worst first.

KEY TERMS

impact — how much damage if a risk occurs.

likelihood — how probable the risk is.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

B In incident response, it is important to first prevent the spread of damage before removing the cause.

C The size of a risk can be determined by impact alone.

D The final stage of incident response is prevention of recurrence.

(2) For each of a–d, indicate which stage of the six stages of incident response (A–D) it belongs to.

a Establishing response procedures in advance, in preparation for incidents

b Disconnecting a malware-infected computer from the network

c Removing the malware infection

d Analyzing the cause and taking measures so that the same incident does not occur again

[Choices] **A** Preparation **B** Containment **C** Eradication **D** Prevention of recurrence

✓ Solution

(1)

A Correct definition of a security incident. Therefore, ○.

B Containment (preventing the spread of damage) is performed before eradication (removing the cause). Therefore, ○.

C The size of a risk is determined by both impact and likelihood. Therefore, ×.

D The final stage of the six stages is prevention of recurrence. Therefore, ○.

(2)

a: Preparing in advance is preparation. **A**

b: Disconnecting from the network to prevent the spread of damage is containment. **B**

c: Removing malware is eradication. **C**

d: Cause analysis and countermeasures are prevention of recurrence. **D**

EXAM-STYLE QUESTION

Explain why the six stages of incident response must be carried out in order, using recovery and eradication as an example. **[6] Refer to:** what happens if the cause is not removed.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the term for a situation in which actual damage occurs due to an information security problem?
2. Name the six stages of incident response in order.
3. What are the two factors used to determine the size of a risk?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate example of a security incident.

A Data was encrypted by a ransomware infection.

B Customer information was leaked to the outside through unauthorized access.

C A software update was performed to add new features.

D A web service became unavailable due to an attack on the server.

(2) From the following options (A–D), choose the one that most appropriately describes the concept of risk assessment.

A It is enough to address only risks with large impact.

B It is enough to address only risks with high likelihood.

C Risks with both high impact and high likelihood should be addressed first.

D All risks must be addressed simultaneously.

⚙ Think as an Engineer — Investigate & Decide

A school's student-records system is hit by ransomware: files are encrypted and staff cannot log in. You lead the response.

☐ **Restore from backup immediately**

☐ **Follow the six-stage response in order**

1 · Respond from here. This incident has just been **detected** — files are encrypted and staff are locked out. Starting from where the school is **now**, list the stages still to carry out, and say what each one means for this specific incident.

2 · Assess the risks. For two risks the school faces — (i) another ransomware attack, (ii) a brief power cut — estimate impact and likelihood, and say which to address first.

3 · Decide. Recommend the first three actions the school should take, and give one reason why order matters. **In pairs,** compare your orders and agree together on the single most important first step.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Give evidence for your answer.

Stuck? If you recover before you eradicate the cause, the malware is still there — and the same incident can happen again.

🌍 In a New Context

An online shop has just **detected** that customer data has been leaked. Using the six-stage response and risk assessment, name the next two stages it should carry out and why, and identify one risk it should prioritize afterward.

❓ Lesson Question — Answered

The lesson asked how to respond to an incident in order, and how to prioritize risks. A security incident is any situation where real damage occurs. The response follows six stages in a fixed order — preparation, detection, containment, eradication, recovery, and prevention of recurrence — because acting out of order (for example, recovering before removing the cause) lets the same incident return. To decide what to guard against first, an organization assesses each risk by its size = impact × likelihood, and addresses high-impact, high-likelihood risks first, since no team can fix everything at once. Responding in order and prioritizing by risk together keep damage small and prevent repeats.

KEY TERMS

containment — limit the spread of damage.

eradication — remove the cause.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

A situation in which actual damage occurs due to an information security problem is called (a). When (a) occurs, (b), which prevents the spread of damage, is performed first, followed by (c), which removes the cause of the incident. After that, (d), which returns the system to its normal state, is performed, and finally (e), in which the cause is analyzed to prevent recurrence, is performed.

⭐ REFLECT & CHALLENGE

Reflect: Which stage of incident response do you think is hardest to do well under pressure, and why? **Challenge:** Suggest one thing an organization could prepare in advance (stage 1) that would make the other phases faster.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1. Fill in the blanks (a)–(e).

3 Answer the following questions.

(1) Reorder the risks (a–d) from highest priority to lowest priority based on the concept of risk assessment (impact × likelihood).

a Impact: large (3), Likelihood: high (3)

b Impact: small (1), Likelihood: high (3)

c Impact: large (3), Likelihood: medium (2)

d Impact: small (1), Likelihood: low (1)

(2) From the following options (A–D), choose the one that most appropriately describes the order of incident response.

A Detection → Eradication → Containment → Recovery

B Preparation → Detection → Containment → Eradication → Recovery → Prevention of recurrence

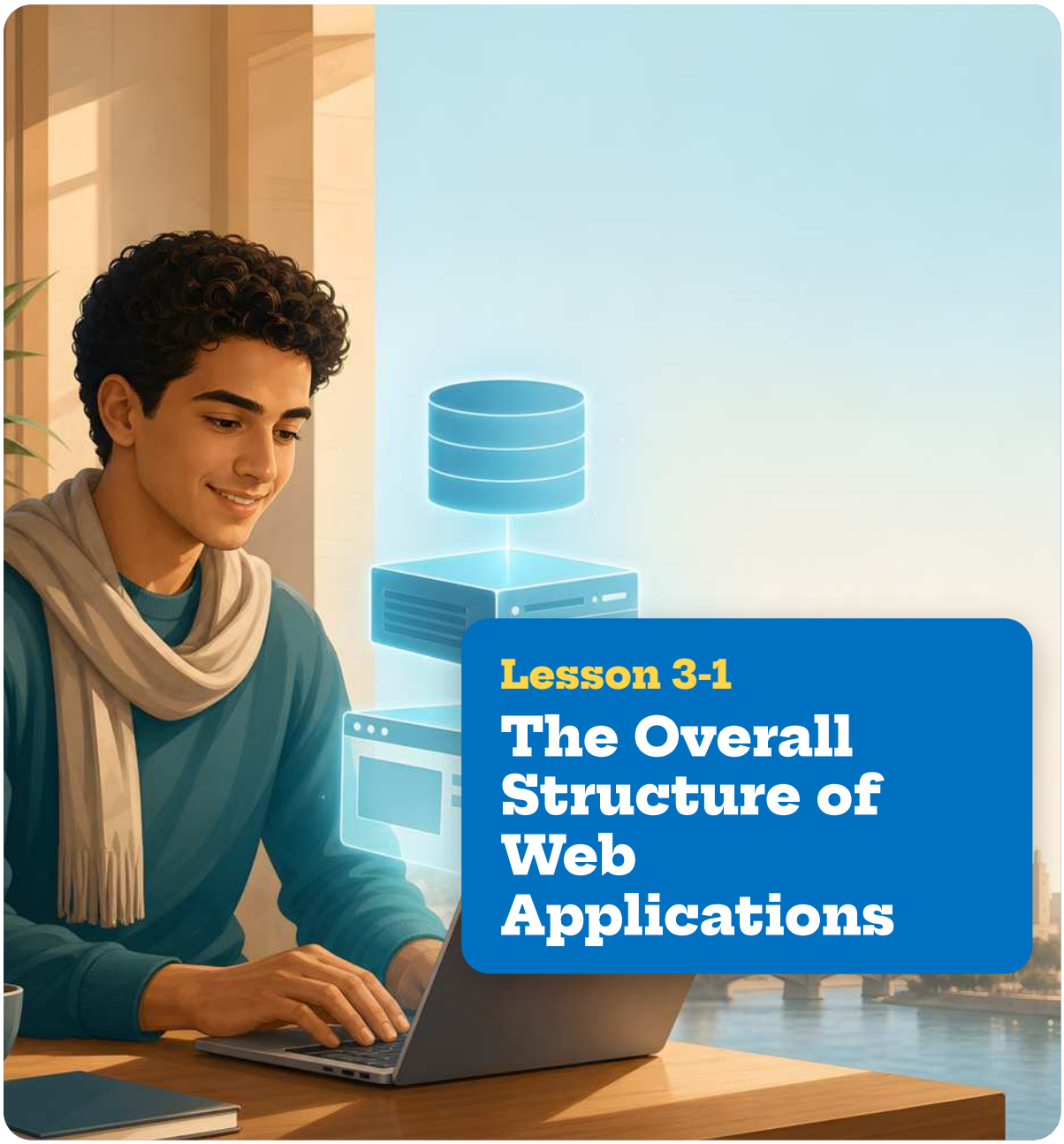
C Detection → Recovery → Containment → Prevention of recurrence

D Containment → Detection → Eradication → Preparation

⭐ Key takeaway

Remember:
Follow the six stages in order (prepare → detect → contain → eradicate → recover → prevent), and prioritize risks by impact × likelihood.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 3-1

The Overall Structure of Web Applications

Learning Objectives

- Explain what a web application is and describe the roles of its three tiers: frontend, backend, and database.
- Trace how the three tiers cooperate to turn a user's action into a result, using familiar services as examples.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Every time you search on YouTube, buy on Amazon, or open Gmail, you are using a web application — a program you use through your browser, with nothing to install. Behind the simple screen, three tiers work together: the page you interact with, the hidden server that processes the request, and the database that holds all the data. Understanding these three tiers explains how almost every online service is built.

? Guiding question: How is a web application structured, and how do its three tiers work together to respond to what a user does?

Key Fact:
A web application is built from three tiers — frontend, backend, and database — that cooperate to turn a user's action into a result.

Key Concepts:

- web application
- frontend
- backend
- database
- three-tier architecture

KEY TERMS
web application — an application used through a web browser, with no dedicated software to install.

Point!

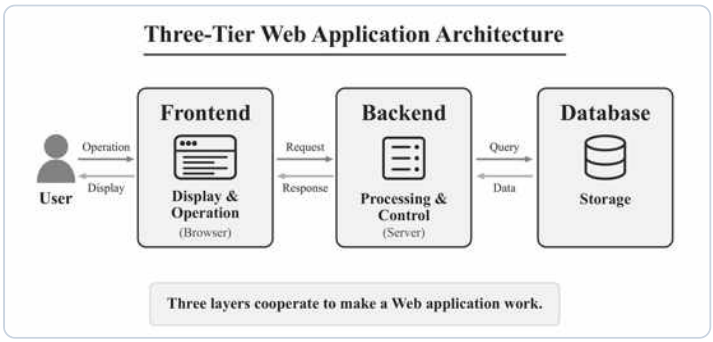
1. What is a Web Application?

(1) **(Web application)**…an application used through a web browser. It can be used as long as you have a browser, without needing to install dedicated software.
e.g., YouTube, Amazon, Gmail, Google Maps

2. The Three Tiers of a Web Application

(1) A web application is broadly composed of the following three tiers.

Tier	Role	Examples
Frontend	The screen component that users see directly. It runs on the browser and accepts user input.	Search screen, product listing pages, input forms, etc.
Backend	The server-side processing that users do not see. It processes data and makes decisions.	Searching for products that match search criteria, login authentication, etc.
Database	A system that organizes, stores, and manages data. It retrieves data when needed and records new data.	Product information, user information, order history, etc.



Three-Tier Web Application Architecture

The data tier stores and serves the records.

Three tiers: presentation, logic, and data.

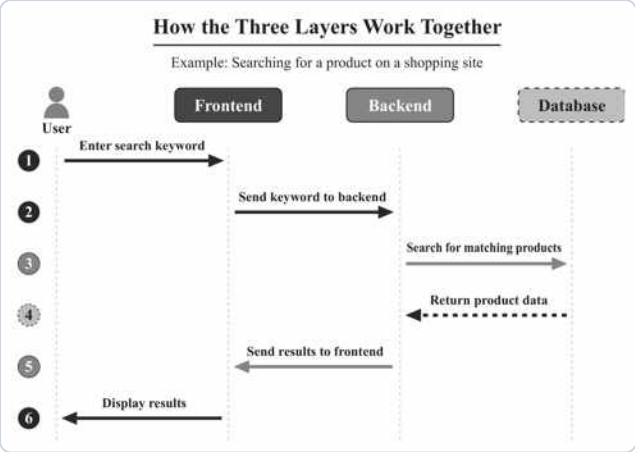
3. The Flow of How the Three Tiers Cooperate

(1) When a user operates a web application, the three tiers cooperate to perform the processing to make it work.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

e.g., Searching for a product on an online shopping site

The frontend receives the search keywords entered by the user.
The frontend sends the search keywords to the backend.
The backend searches the database for products that match the criteria.
The database returns the data of matching products to the backend.
The backend organizes the search results and returns them to the frontend.
The frontend displays the search results on the user's screen.



How the Three Layers Work Together

4. Familiar Web Application Configurations

(1) All familiar web applications also operate on this three-tier architecture.

Service	Frontend (display and interaction)	Backend (processing and control)	Database (storage and management)
Video sharing site	Display of video playback and search screens; reception of user operations	Selection of recommended videos; aggregation of view counts	Storage of video data, user information, and viewing history
SNS	Display of timelines and post screens; sending of input content	Determining the audience for posts; control of push notifications	Storage of post data, follow relationships, and messages
E-commerce site	Display of product listings and cart screens; reception of order operations	Stock checking; execution of payment systems	Storage of product information, order history, and customer information

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A The frontend is the screen component that users see directly.

B The backend is the part that runs on the browser.

C A database is a system that organizes, stores, and manages data.

D A web application operates with the frontend alone.

PAUSE & THINK

When you type in a search box and see results, which tier collected your input, and which tier found the answer?

A request flows down the tiers and a reply returns.

KEY TERMS

frontend — the screen users see and interact with.
backend — the hidden server-side processing.
database — where data is stored and managed.

PAUSE & THINK

In the table, every service uses the same three tiers — what changes from one service to the next?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a The part responsible for server-side processing that users do not see

b The screen component that users interact with directly

c A system that organizes, stores, and manages data

[Choices] **A** Frontend **B** Backend **C** Database

✓ Solution

(1)

A The frontend is the screen component that users see directly. Therefore, ○.

B The backend is the part that runs on the server, not on the browser. Therefore, ×.

C Correct definition of a database. Therefore, ○.

D A web application operates with three tiers: frontend, backend, and database. Therefore, ×.

(2)

a: Server-side processing is the backend. **B**

b: The screen that users interact with is the frontend. **A**

c: Storing and managing data is the database. **C**

Try

1 Answer the following questions.

1. What is the term for an application used through a web browser?
2. What is the term for the screen component that users see directly?
3. What is the term for the part responsible for server-side processing that users do not see?
4. What is the term for a system that organizes, stores, and manages data?

2 Answer the following questions.

(1) When searching for a product on an online shopping site, which of the following is responsible for searching through data for products that match the search keywords: **A** Frontend, **B** Backend, **C** Database?

EXAM-STYLE QUESTION

Explain how the three tiers of a web application cooperate when a user posts a photo on a social media app. **[6] Refer to:** the role of the frontend, backend, and database.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) For each of a–f, indicate which tier of an e-commerce site it belongs to: **A** Frontend, **B** Backend, **C** Database.

a Product listing page

b Stock checking processing

c Storage of customer information

d Cart screen

e Payment processing

f Storage of order history

⚙ Think as an Engineer — Investigate & Decide

A clinic wants a web app where patients book appointments, check which time slots are free, and see their past visits. You plan its three tiers.

☐ **Put everything in one program**

☐ **Separate the frontend, backend, and database**

1 · Assign the work. For three features — the booking form, checking which time slots are free, and storing past visits — say which tier (frontend, backend, or database) should handle each, and why.

2 · Diagnose a fault. The booking page opens normally, but no free slots appear on screen. Which tier is clearly working, and which one has probably failed? Explain your reasoning.

3 · Decide. Recommend whether the clinic should separate its app into three tiers, and give two reasons. **In pairs,** compare your assignments and agree together on where checking free time slots belongs.

Give evidence for your answer.

Stuck? The frontend shows and collects; the backend decides and processes; the database stores and retrieves.

🌍 In a New Context

A weather app lets users type a city and see the forecast. Using the three-tier model, name which tier does each job — taking the city name, looking up the forecast data, and storing past forecasts — and identify one job the frontend cannot do on its own.

KEY TERMS

three-tier architecture — frontend + backend + database cooperating.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

❓ Lesson Question — Answered

The lesson asked how a web application is structured and how its tiers cooperate. A web application runs in a browser and is built from three tiers. The frontend is the screen the user sees and interacts with; the backend is the hidden server-side processing that makes decisions; and the database organizes, stores, and manages the data. When a user acts — for example, searching for a product — the frontend collects the input and passes it to the backend, the backend queries the database, the database returns the data, and the backend sends the organized result back to the frontend to display. Overall, a web application works not through one part but through three tiers cooperating, each with a clear role.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

A web application is broadly composed of three tiers. The screen component that users see directly is called the (a). The part responsible for server-side processing that users do not see is called the (b). A system that organizes, stores, and manages data is called the (c). In response to user actions, these three tiers cooperate to perform processing.

1. Fill in the blanks (a)–(c).

3 Answer the following questions.

(1) For each of a–f, indicate which tier of an SNS it belongs to: **A** Frontend, **B** Backend, **C** Database.

a Timeline screen
b Processing that determines the audience for a post
c Storage of post data
d Post screen
e Notification processing
f Storage of follow relationships

⭐ REFLECT & CHALLENGE

Reflect: Which tier do you think fails most noticeably to a user when something goes wrong, and why? **Challenge:** Name one online service you use and describe what its frontend, backend, and database each do.

⭐ Key takeaway

Remember:
Frontend shows and collects, backend decides and processes, database stores and retrieves — three tiers cooperating make a web application work.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) From the following options (A–D), choose the one that most appropriately describes the three-tier architecture of a web application.

A The frontend has the role of storing and managing data.

B The backend is the part that users interact with directly on the browser.

C In response to user actions, the frontend, backend, and database cooperate to perform processing.

D The database has the role of displaying screens to users.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 3-2

Web Application Communication Methods

Learning Objectives

- Explain the client and server model and the difference between HTTP and HTTPS.
- Describe how HTTP requests and responses work, and explain the role of an API in exchanging data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

When you click a link or press “search,” your device does not hold the whole website — it asks a distant computer for it. Your browser is the client that makes the request; a server somewhere answers with a response. This exchange, governed by shared rules called protocols, is how every web application communicates. Knowing how requests, responses, and APIs work explains what happens between your click and the page you see.

? Guiding question: How do the client and server communicate to make a web application work, and what rules and messages do they use?

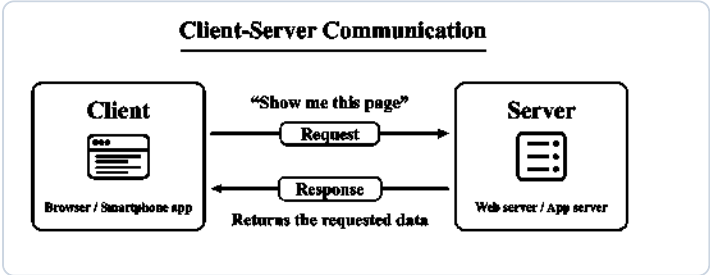
Point!

1. Client and Server

(1) **(Client)**…the side that requests data or processing from the server. Web browsers and smartphone apps fall into this category.

(2) **(Server)**…the side that receives requests from clients, processes them, and returns responses.

(3) A web application operates by the client sending a request and the server returning a response.



Client-Server Communication

2. HTTP and HTTPS

(1) **(HTTP (HyperText Transfer Protocol))**…a communication protocol for exchanging data between clients (browsers and apps) and servers.

(2) **(HTTPS (HTTP Secure))**…a communication protocol that adds TLS encryption to HTTP. Because communication content is encrypted, the risk of its being read by third parties is reduced.

(3) Websites whose URLs begin with “https://” use HTTPS communication.

3. HTTP Requests and Responses

(1) Requests sent from the client to the server include information called an HTTP method that indicates “what kind of operation is desired.”

Key Fact:
Web applications work by a client sending requests and a server returning responses, using protocols (HTTP/HTTPS) and often APIs that exchange data in formats like JSON.

Key Concepts:

- client / server
- HTTP / HTTPS
- HTTP method
- status code
- API / JSON

KEY TERMS
client — the side that requests data or processing.
server — the side that processes requests and returns responses.

Many clients request; one server responds.

PAUSE & THINK

Why should a login page use HTTPS rather than plain HTTP?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Method	Role	Examples
GET	Retrieve data from the server	Displaying a web page, retrieving search results
POST	Send data to the server	Sending login information, submitting form contents

(2) Responses returned from the server to the client include a number indicating the result of processing, called a status code. Some typical examples are as follows.

Status code	Meaning
200	Success (the request was processed successfully)
404	Page not found (the specified URL does not exist)
500	Server error (a problem occurred on the server side)

4. API (Application Programming Interface)

(1) **(API)**…a system for exchanging data between different pieces of software.

(2) In web applications, the frontend often sends requests to the backend through an API and receives data.

(3) Data exchanged through an API is often described in a format called **(JSON (JavaScript Object Notation))**. JSON is a data format that is both easy for humans to read and for computers to process.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A The client is the side that requests data or processing from the server, and the server is the side that returns the response.

B HTTPS is a communication protocol that adds encryption to HTTP.

C The GET method is for sending data to the server.

D Status code 404 indicates that the page is not found.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A communication protocol for exchanging data between web browsers and servers

b A system for exchanging data between different pieces of software

c A communication protocol that adds encryption to HTTP

[Choices] **A** HTTP **B** API **C** HTTPS

KEY TERMS

HTTP method — says what is wanted (GET, POST).
status code — reports the result (200, 404, 500).

One HTTP request, one matching response.

An API is a fixed contract for exchanging data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✓ Solution

(1)

A Correct description of the roles of client and server. Therefore, ○.

B HTTPS adds SSL/TLS encryption to HTTP. Therefore, ○.

C The GET method is for retrieving data from the server and the POST method is for sending data. Therefore, ×.

D 404 is a status code indicating that the page is not found. Therefore, ○.

(2)

a: The web communication protocol is HTTP. **A**

b: Data exchange between pieces of software is API. **B**

c: HTTP with added encryption is HTTPS. **C**

Try

1 Answer the following questions.

1. What is the term for the side that requests data or processing from the server (e.g., a web browser)?
2. What is the term for the side that receives requests from clients, processes them, and returns responses?
3. What is the term for the communication protocol that adds encryption to HTTP?
4. What is the term for the HTTP method for retrieving data from the server?
5. What is the term for the system for exchanging data between different pieces of software?

2 Answer the following questions.

(1) From the following options (A–D), choose the combination of status code and meaning that is NOT appropriate.

A 200 — The request was processed successfully

B 404 — Page not found

C 500 — The request was processed successfully

D 500 — A problem occurred on the server side

EXAM-STYLE QUESTION

Explain what happens, step by step, when a browser requests a web page and the server responds. **[6] Refer to:** client, request, HTTP method, response, status code.

PAUSE & THINK

If a browser receives status code 404, what should it tell the user — and whose side had the problem?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) From the following options (A–D), choose the one that most appropriately describes an API.

A A technology for adjusting the appearance of web pages

B A system for exchanging data between different pieces of software

C A technology for encrypting data and sending it safely

D An HTTP method for retrieving data from the server

⚙ Think as an Engineer — Investigate & Decide

A school in your city wants a web app where students log in, view their timetable, and download a PDF. You plan how the client and server communicate.

☐ **Send everything over plain HTTP**

☐ **Use HTTPS and the right method for each action**

1 · Choose the method. For three actions — loading a student's grade history, submitting the login form, and requesting today's schedule from another school system — say whether GET or POST fits, and for the last say whether an API is involved.

2 · Choose the protocol. Decide whether the login should use HTTP or HTTPS, and give the reason.

3 · Decide. Recommend the communication design and give two reasons. **In pairs,** compare your method choices and agree together on which action must use POST.

Give evidence for your answer.

Stuck? GET retrieves and shows; POST sends data such as a login; HTTPS encrypts so passwords are not intercepted.

🌍 In a New Context

A weather service offers an API that returns today's forecast when a client sends a request. Using what you learned, name which HTTP method the client would use to retrieve the forecast, why the service should use HTTPS, and one advantage of receiving the data as JSON.

KEY TERMS

API — lets different software exchange data.

JSON — a readable data format.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

❓ Lesson Question — Answered

The lesson asked how the client and server communicate and what rules and messages they use. A web application works when a client — a browser or app — sends a request and a server processes it and returns a response. They communicate through a protocol: HTTP carries the messages, and HTTPS adds SSL/TLS encryption so the content is not easily intercepted. Each request carries an HTTP method that says what is wanted — GET to retrieve, POST to send — and each response carries a status code that reports the result, such as 200 for success or 404 for not found. Often the frontend and backend exchange data through an API, frequently in the JSON format, which is easy for both people and computers to read. Overall, communication in a web application is an organized exchange of requests and responses, shaped by protocols, methods, status codes, and APIs.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

A web application operates by communication between a (a), which requests data or processing from the server, and a (b), which receives the request, processes it, and returns a response. This communication uses a protocol called (c). The protocol that adds SSL/TLS encryption to (c) is called (d).

1. Fill in the blanks (a)–(d).

3 Answer the following questions.

(1) For each of operations (a–d), indicate which HTTP method is used: **A** GET, **B** POST.

a Displaying a web page

b Entering an ID and password into a login form and submitting them

c Displaying search results

d Submitting the contents of a contact form

(2) From the following options (A–D), choose the one that most appropriately describes HTTPS.

A A method for retrieving data from the server

B A system for exchanging data between different pieces of software

C A communication protocol that adds SSL/TLS encryption to HTTP

D A status code indicating that a problem has occurred on the server side

⭐ REFLECT & CHALLENGE

Reflect: When a page fails to load, which status code would help you most to understand why, and what would it tell you?
Challenge: Name one app you use and describe one request it sends and the response it expects.

⭐ Key takeaway

Remember:
Client requests, server responds — over HTTP/HTTPS, with a method (GET/POST), a status code, and often an API returning JSON.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 3-3

Fundamentals of Frontend Technology

Learning Objectives

- Explain the distinct roles of HTML, CSS, and JavaScript in building a web page.
- Describe how semantic HTML, responsive design, and frameworks improve a modern frontend.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Open any web page and you are seeing three technologies at work together. HTML builds the structure — the headings, paragraphs, and links, like the frame of a building. CSS adds the appearance — colors, layout, and fonts, like a building's interior. JavaScript adds behavior — search boxes that respond, buttons that count, screens that update. A modern frontend goes further: it makes pages meaningful, adaptable to every screen, and faster to build.

? Guiding question: How do HTML, CSS, and JavaScript work together to build a web page, and what makes a modern frontend meaningful, adaptable, and efficient to build?

Key Fact:
A web page is built from HTML (structure), CSS (appearance), and JavaScript (behavior); a modern frontend adds semantic HTML, responsive design, and frameworks.

Key Concepts:

- HTML / CSS / JavaScript
- semantic HTML
- accessibility / SEO
- responsive design
- framework

KEY TERMS
HTML — structure.
CSS — appearance.
JavaScript — behavior.

HTML structures, CSS styles, JS adds behavior.

Point!

1. The Roles of HTML, CSS, and JavaScript

(1) Web pages are mainly based on the following three technologies.

Technology	Role	Analogy
HTML	Defines the structure (skeleton) of a web page. Describes elements such as headings, paragraphs, images, and links.	The frame of a building
CSS	Specifies the appearance of a web page (colors, layout, size, fonts, etc.).	The interior and exterior of a building
JavaScript	Adds behavior and interactivity to a web page. Activates responses when buttons are pressed and updates only part of the screen.	The electrical and plumbing systems of a building

Roles of HTML, CSS, and JavaScript

(2) These three technologies each have independent roles, and a web page is created using all of them in combination.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

2. Semantic HTML

(1) **(Semantic HTML)**…the concept of describing each part of a web page using element names (tags) that indicate their meaning.
e.g., Use <header> at the top of the page, <nav> for navigation, <main> for the body, and <footer> at the bottom of the page.

(2) Benefits of using semantic HTML

Improved **(accessibility)**…screen reader software can correctly understand the structure of the page. Information becomes easier to convey to users with visual impairments.

Improved **(SEO (Search Engine Optimization))**…search engines can more easily understand the content of a web page correctly, making it more likely to be displayed appropriately in search results.

3. Responsive Design

(1) **(Responsive design)**…a design concept in which the layout is automatically adjusted according to different screen sizes such as PCs, smartphones, and tablets.

(2) With a single web page, a display that is easy to view from any device can be achieved.

(3) As of 2026, websites are frequently accessed from smartphones, and support for responsive design is widely required. Against this backdrop, the **(mobile-first)** concept—prioritizing small screens such as smartphones from the design stage—has also become widespread.

4. Frameworks

(1) **(Framework)**…a structure that provides commonly used functions and components in advance, in order to make web application development more efficient.

(2) By using a framework, developers do not need to write all code from scratch and can develop more efficiently.

(3) Major frontend frameworks

Framework name	Characteristics
React	Developed by Meta (formerly Facebook). It can efficiently update only part of the screen and is suited to large-scale web applications.
Vue	Has a simple structure and allows web applications to be built incrementally from small features.
Next.js	In addition to React's features, includes mechanisms to speed up page display.

PAUSE & THINK

On a page you use often, which parts are structure (HTML), which are style (CSS), and which react to you (JavaScript)?

Responsive design re-flows to each screen size.

KEY TERMS
responsive design — layout adapts to screen size.
framework — ready-made functions for efficient development.

Frameworks assemble pages from ready-made parts.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A HTML is a technology that defines the structure (skeleton) of a web page.

B CSS is a technology that adds behavior to a web page.

C Responsive design is a concept in which the layout is automatically adjusted according to different screen sizes.

D Using a framework eliminates the need to write all code from scratch and allows efficient development.

(2) Match each role (a–c) with the most appropriate technology from the choices below (A–C).

a Defines the structure of a web page

b Specifies the appearance of a web page

c Adds behavior and interactivity to a web page

[Choices] **A** HTML **B** CSS **C** JavaScript

✓ Solution

(1)

A HTML defines the structure of a web page. Therefore, ○.

B CSS is a technology that specifies the appearance of a web page. Adding behavior is JavaScript. Therefore, ×.

C Correct definition of responsive design. Therefore, ○.

D Correct description of the benefit of frameworks. Therefore, ○.

(2)

a: Defining structure is HTML. **A**

b: Specifying appearance is CSS. **B**

c: Adding behavior is JavaScript. **C**

PAUSE & THINK

Why does a screen reader understand a page better when it uses <header>, <nav>, and <main> instead of plain boxes?

EXAM-STYLE QUESTION

Explain how HTML, CSS, and JavaScript each contribute to a library page that lists books, is styled, and filters the results as the user types. **[6] Refer to:** structure, appearance, behavior.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the four-letter abbreviation for the technology that defines the structure (skeleton) of a web page?
2. What is the three-letter abbreviation for the technology that specifies the appearance of a web page (colors, layout, size, etc.)?
3. What is the term for the technology that adds behavior and interactivity to a web page?
4. What is the term for the design concept in which the layout is automatically adjusted according to different screen sizes?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate benefit of using semantic HTML.

A Screen reader software can correctly understand the structure of the page.

B Search engines can more easily understand the content of a web page.

C The display speed of a web page necessarily becomes faster.

D Information becomes easier to convey to users with visual impairments.

(2) From the following options (A–D), choose the one that most appropriately describes a framework.

A A language for defining the structure of a web page

B A structure for making web application development more efficient

C A technology for encrypting data and sending it safely

D A design concept for supporting different screen sizes

PAUSE & THINK

Which of the benefits below is about how the page looks, and which is about how well others can find and read it?

Think as an Engineer — Investigate & Decide

A student club wants a simple web page that lists its events, is styled in the club's colors, and has a button showing how many people have signed up.

☐ **Write everything in one file**

☐ **Use HTML, CSS, and JavaScript for their proper roles**

1 · Observe first. Open one web page you use often and note two things on it that change when you click or type. Then, for three features of the club page — the list of events, the club's colors and layout, and the live sign-up counter — say which technology (HTML, CSS, or JavaScript) does each, and why.

2 · Make it reach everyone. Name one change that would make the page easier to read on a phone (responsive design), and one semantic-HTML choice that would help a screen reader.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 · Decide. Recommend whether the club should separate HTML, CSS, and JavaScript, and give two reasons. **In pairs,** compare your role assignments and agree together on what the sign-up counter needs.

Give evidence for your answer.

Stuck? HTML holds the content; CSS styles it; JavaScript makes it react and update.

🌍 In a New Context

A news site must look good on phones and be found easily by search engines. Using what you learned, name one responsive-design choice and one semantic-HTML choice it should make, and say what each one improves.

KEY TERMS

accessibility — usable by people with impairments.
SEO — being found by search engines.

❓ Lesson Question — Answered

The lesson asked how HTML, CSS, and JavaScript work together and what makes a modern frontend strong. A web page is built from three technologies with independent roles: HTML defines the structure, CSS specifies the appearance, and JavaScript adds behavior and interactivity — combined, they create the page. A modern frontend goes further: semantic HTML uses meaningful tags so screen readers and search engines understand the page (improving accessibility and SEO); responsive design adjusts the layout to any screen size, with a mobile-first approach; and frameworks such as React, Vue, and Next.js provide ready-made functions so developers build efficiently without writing everything from scratch. A modern web page is therefore not one technology but a combination — structured, styled, interactive, meaningful, adaptable, and efficient to build.

Exercise

⭐ REFLECT & CHALLENGE

Reflect: Of structure, style, and behavior, which do you notice most when a web page is badly made, and why? **Challenge:** Name one website you use and describe one thing HTML, CSS, and JavaScript each do on it.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Web pages are mainly composed of three technologies: (a), which defines the structure of the page; (b), which specifies the appearance; and (c), which adds behavior and interactivity. The concept of adjusting the layout according to different screen sizes such as PCs, smartphones, and tablets is called (d).

1. Fill in the blanks (a)–(d).

3 Answer the following questions.

(1) For each of tasks (a–f), indicate which technology is used to achieve it: **A** HTML, **B** CSS, **C** JavaScript.

a Describing the structure of headings and paragraphs

b Changing the color and size of text

c Updating part of the screen when a button is pressed

d Placing images and links

e Setting the background color of the entire page

f Validating the contents entered into a form

(2) From the following options (A–D), choose the one that most appropriately describes the reason responsive design is needed.

A To improve the display speed of websites

B Because websites are accessed from various devices such as PCs, smartphones, and tablets

C To raise rankings in search engines

D To add behavior to web pages

Key takeaway

Remember:
HTML structures, CSS styles, JavaScript behaves — and semantic HTML, responsive design, and frameworks make a modern frontend meaningful, adaptable, and efficient.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 4-1

Types and Characteristics of Media

Learning Objectives

- Compare the characteristics of different media, and explain the strengths and constraints of websites.
- Choose suitable media for a given purpose and audience, and justify the choice.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Before building a website, it helps to ask: why a website at all? TV reaches millions at once but only one way. Radio can be received almost anywhere but conveys sound only. A newspaper lasts on paper without the Internet. A website combines text, images, and video, updates at any time, and can even respond through forms — while SNS (social networking services) spreads news the fastest of all. Each medium has its own strengths and constraints, so the smart choice depends on the message and the audience.

? Guiding question: What are the strengths and constraints of different media, and how do you choose the right one for a purpose and audience?

Key Fact:
Each medium has different strengths and constraints; choosing the right medium — often several combined — depends on the purpose and the target audience.

Key Concepts:

- TV / radio / newspaper
- website / SNS
- one-way / two-way
- strengths / constraints
- combining media

KEY TERMS
one-way — from sender to receiver only.
two-way — users can respond and share.

Point!

1. Comparison of the Characteristics of Different Media

(1) Each medium for conveying information has different characteristics. When planning a website, it is important to understand the differences from other media.

Medium	Information conveyed	Characteristics
TV	Video + audio	Reaches many people simultaneously. Appeals to sight and hearing. One-way, from sender to receiver.
Radio	Audio	Appeals only to hearing. Can be received almost anywhere and alongside other tasks. One-way.
Newspaper	Text + images	Remains physically as a printed medium. Can be read without an Internet connection. One-way.
Website	Text + images + video + audio	Diverse information can be combined. Can be updated at any time. Basically one-way, but two-way interaction is also possible through forms, comments, etc.
SNS	Text + images + video + audio	Highly effective at spreading information rapidly. Users can post and share information. Two-way.

2. Strengths and Constraints of Websites

(1) Strengths of websites

① Diverse information such as text, images, and video can be combined and presented.

② Information can be updated at any time.

③ Can be accessed from almost anywhere in the world.

④ Two-way interaction with users is possible through forms, comments, etc.

⑤ Can be used with just a browser; no dedicated app installation is required.

Traditional and digital media, side by side.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) Constraints of websites

① Normally requires an Internet connection (unless the page is cached).

② Information does not reach users unless they actively access it.

③ Display and loading speed may differ depending on the user's device and network environment.

A website combines strengths one-way media lack.

PAUSE & THINK

Which medium would you choose to reach the most people at once — and which to let them respond to you?

3. Choosing Media According to Purpose

(1) When conveying information, it is important to choose the most suitable medium according to the purpose and target audience.

Purpose	Examples of suitable media	Reason
Inform many people at once	TV, SNS ads	Wide reach
Provide detailed information	Website	No limit on information volume; can be viewed at any time
Deliver information directly to specific people	Email, direct messages	Can be delivered individually
Spread by word of mouth	SNS	Users can voluntarily share and spread information

Match the medium to the purpose and audience.

(2) In practice, multiple media are often used in combination.
e.g., A TV commercial directs viewers to a website, and the website encourages spreading information through SNS.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A website can present diverse information such as text, images, and video in combination.

B TV is a medium that allows two-way interaction with users.

C SNS is a medium highly effective at spreading information rapidly.

D A website normally can be viewed even without an Internet connection.

(2) Match each purpose (a–c) with the most suitable medium from the choices below (A–C).

a Make detailed information viewable at any time

b Have users spread information by word of mouth

c Inform many people at once with video

[Choices] **A** TV **B** Website **C** SNS

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✓ Solution

(1)

A Correct as a strength of websites. Therefore, ○.

B TV is a one-way medium from sender to receiver. Therefore, ×.

C Correct description of the characteristics of SNS. Therefore, ○.

D A website normally requires an Internet connection. Therefore, ×.

(2)

a: A website allows detailed information to be viewed at any time. **B**

b: SNS spreads easily by word of mouth. **C**

c: TV reaches many people with video. **A**

Try

1 Answer the following questions.

1. Name two strengths of websites.
2. Name two constraints of websites.

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate characteristic of websites.

A Information can be updated at any time.

B Information automatically reaches users without them actively accessing it.

C Text, images, and video can be combined and presented.

D Two-way interaction is possible through forms and comments.

(2) For each situation (a–d), indicate which medium is most suitable: **A** TV, **B** Website, **C** SNS, **D** Email.

a Make the detailed specifications and usage of a new product available at any time.

b Have users spread information about a new product by word of mouth.

c Deliver video of a new product to many people at once.

d Send campaign information directly to existing customers.

⚙ Think as an Engineer — Investigate & Decide

A public library in your city wants to tell people about a new free reading program. You advise on which media to use.

☐ **Use one medium only** ☐ **Combine media by purpose**

1 · Collect data. Survey ten classmates or family members: where do they actually notice local announcements — a printed poster, Facebook,

EXAM-STYLE QUESTION

Explain why an organization might use a website, TV, and SNS together rather than only one, to promote a new service. **[6] Refer to:** reach, depth, and spreading by word of mouth.

PAUSE & THINK

For each situation below, ask first: is the goal wide reach, full detail, spreading, or a direct message?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

WhatsApp, or local radio? Record the results in a table and find the most common channel.

2 · Analyze the audience. For each group — a family with no home Internet, a student, and the librarian — note one benefit and one drawback of using a website as the library's main channel.

3 · Decide. Using your survey and your table, recommend a media plan (which media, in what roles) and give two reasons. **In pairs,** compare your plans and agree together on the single best medium for reaching families who may not be online.

Give evidence for your answer.

Stuck? Start from the audience your survey shows is hardest to reach — which medium already reaches them where they are?

🌍 In a New Context

A hospital wants to inform patients about new opening hours and let them ask questions. Name one medium for reaching many people quickly and one that allows two-way interaction, and say why each one fits.

❓ Lesson Question — Answered

The lesson asked about the strengths and constraints of different media and how to choose. Each medium has its own character: TV reaches many at once but one-way; radio is sound only but reaches anywhere; a newspaper lasts on paper without the Internet; a website combines text, images, and video, updates any time, and allows two-way interaction, though it needs an Internet connection and only reaches those who seek it; and SNS spreads information fastest and is two-way. Because no single medium is best for everything, the right choice depends on the purpose and the target audience — wide reach, full detail, direct delivery, or word-of-mouth — and in practice several media are combined so each does what it does best.

KEY TERMS

website strengths — diverse, updatable, global, two-way.

website constraints — needs Internet; users must seek it.

Exercise

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Each medium for conveying information has different characteristics. A website has the strengths of being able to present diverse information such as text, images, and video, and being updatable at any time. On the other hand, it has the constraints that it cannot be viewed without an (a) connection, and that information does not reach users unless they actively (b) it. In actual information dissemination, (c) media are often combined, such as directing viewers from a TV commercial to a website.

1. Fill in the blanks (a)–(c).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the characteristics of media.

A Radio is a medium that conveys information through video and audio.

B Newspapers are media highly effective at spreading information rapidly.

C SNS is a medium that allows users to post and share information.

D TV is a medium that allows two-way interaction with users.

(2) From the following options (A–D), choose the one that most appropriately describes the choice of media.

A Information dissemination should always use only one medium.

B Regardless of purpose or target audience, a website is sufficient.

C It is important to choose the most suitable medium according to the purpose and target audience.

D With SNS, TV and websites are not needed.

REFLECT & CHALLENGE

Reflect: Which medium do you personally trust most for important news, and why?
Challenge: Plan the media mix for announcing a school event on a budget that rules out TV — then name one audience your mix might still miss, and why.

Key takeaway

Remember:
No medium is best for everything — match the medium to the purpose and audience, and combine media so each does what it does best.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 4-2

Information Design and User Experience for Websites

Learning Objectives

- Explain how personas and wireframes are used to plan a website before designing its appearance.
- Apply the four principles of design (contrast, repetition, alignment, proximity) and explain the idea of user-centered design.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

A website that makes sense to its designer can still confuse its visitors. Good design therefore begins with the user, not the creator. Designers first picture a real user (a persona), sketch where information will go before choosing any colors (a wireframe), and follow four simple principles — contrast, repetition, alignment, and proximity — so the page is clear at a glance. This user-first mindset is called user-centered design.

? Guiding question: How do designers plan a website around its users, and what principles make a page clear and easy to use?

Point!

1. Persona

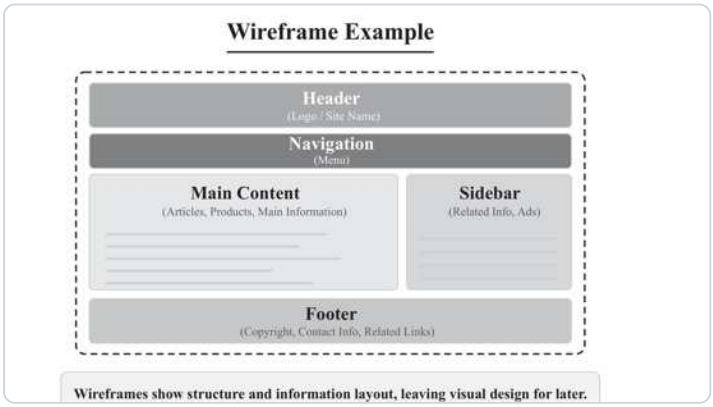
(1) **(Persona)**…a typical user image of those who will use the website, set up as a concrete person. Details such as age, occupation, purpose, and challenges are selected.

(2) By setting up a persona, designers can think specifically about "what this user needs," and the direction of the design becomes less likely to drift.
e.g., Person in their 30s, an office worker, wants to compare products on a smartphone during commute, wants to obtain information in a short time.

2. Wireframe

(1) **(Wireframe)**…a rough blueprint of the layout of a web page. It does not include color or design decoration; it is created to determine the placement and priority of information.

(2) By using a wireframe, "what information to place where" can be considered separately from visual design. If decorations such as colors and fonts are decided first, discussions tend to drift toward personal preferences in appearance, so finalizing the placement and priority of information first makes it easier for stakeholders to reach agreement.



Wireframe Example

Key Fact:
Good web design starts from the user: a persona clarifies who they are, a wireframe fixes what goes where, the CRAP principles make it clear, and user-centered design keeps every choice on the user's side.

Key Concepts:

- persona
- wireframe
- contrast / repetition
- alignment / proximity
- user-centered design

KEY TERMS
persona — a typical user pictured as a concrete person.
wireframe — a layout blueprint before any color or decoration.

A persona captures one target user's goals.

A wireframe grows into a finished design.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3. The Four Principles of Design (CRAP Principles)

(1) Designs that are easy to view and understand follow these four principles.

Principle	Description (purpose)	Specific techniques and examples
(Contrast)	Make the priority of information clear and immediately highlight the most important elements	• Make headings large and bold to differentiate them from body text • Use accent colors for important buttons (e.g., purchase)
(Repetition)	Repeat specific characteristics so that the whole design has consistency, allowing users to navigate without getting lost	• Standardize the color, size, and font of headings on each page • Match the design of bullet point icons
(Alignment)	Align the position of each element to a common baseline, giving order to the entire screen	• Align the left edges of text and images precisely • Make a consistent rule for whether to use center or left alignment
(Proximity)	Place related information close together so that it can be perceived as a single group at a glance	• Group product images, names, and prices together • Leave appropriate spacing (margins) between items

CRAP Design Principles

4. User-Centered Design (UCD)

(1) **(User-Centered Design (UCD))**…a design approach that designs websites and services from the perspective of the user, rather than the convenience of the creator.

(2) In UCD, design begins from “what the user needs” and “in what situation it will be used.” Setting up a persona is one specific way of practicing UCD; by

PAUSE & THINK

Why decide where information goes (wireframe) before choosing colors and fonts?

KEY TERMS

CRAP — contrast, repetition, alignment, proximity.
user-centered design — designing from the user's perspective.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

clarifying the user image, design decisions can be kept consistently aligned with the user's perspective.

(3) Improving usability for users and accessibility for a wide range of people are important goals of UCD.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A persona is a typical user image of those who will use the website, set up as a concrete person.

B A wireframe is a final-version design that includes color and design decoration.

C “Proximity” in the four principles of design means grouping related information close together.

D User-centered design is a design approach that prioritizes the convenience of the creator.

(2) Match each description (a–d) with the most appropriate principle of design from the choices below (A–D).

a Highlight important information and visually indicate priority

b Repeat the same style to give a unified appearance

c Align the positions of elements to give an organized look

d Group related information close together

[Choices] **A** Contrast **B** Repetition **C** Alignment **D** Proximity

✓ Solution

(1)

A Correct definition of a persona. Therefore, ○.

B A wireframe is a blueprint for determining the placement and priority of information, without including color or design decoration. Therefore, ×.

C Correct definition of proximity. Therefore, ○.

D User-centered design is a design approach from the perspective of the user. Therefore, ×.

(2)

a: Highlighting information is contrast. **A**

b: Creating a unified appearance is repetition. **B**

c: Aligning positions is alignment. **C**

d: Grouping related information is proximity. **D**

User-centered design loops around the real user.

PAUSE & THINK

Which principle groups related items together, and which makes the most important item stand out?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the term for a typical user image of those who will use the website, set up as a concrete person?
2. What is the term for a blueprint that shows the placement and priority of information on a web page?
3. What are the four principles of design called when expressed by their English initials?
4. What is the term for the design approach that designs websites and services from the perspective of the user? Also give its three-letter abbreviation.

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate piece of information to include in setting up a persona.

A The user's age and occupation

B The user's purpose for using the website

C The performance of the website's server

D The challenges the user faces

(2) For each of design improvements (a–d), indicate which of the four principles of design (A–D) it corresponds to.

a Making headings large and bold to distinguish them from body text

b Grouping product names, prices, and photos close together

c Aligning the left edges of text and images

d Standardizing the font and color of headings across all pages

A Contrast **B** Repetition **C** Alignment **D** Proximity

Think as an Engineer — Investigate & Decide

A community clinic in your city wants a simple website so patients can find opening hours, services, and directions. You lead the design.

☐ **Start by choosing colors and fonts** ☐ **Start from the user, then design**

1 · Collect data. Ask about eight to ten classmates: "If you opened a community clinic's website, what is the one thing you would look for first?" Record the answers in a tally table.

2 · Analyze your users. Group the answers into user types. Include at least one user who is not a typical smartphone visitor — for example an elderly

EXAM-STYLE QUESTION

Explain why a designer creates a persona and a wireframe before choosing colors and fonts. **[6] Refer to:** the user's needs and reaching agreement with stakeholders.

PAUSE & THINK

For each improvement below, ask: does it highlight, unify, order, or group?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

patient with low vision, or someone with no home Internet using a library computer. Write what each type needs most.

3 · Decide from your data. Using your tally, decide the priority order of the four home-page blocks and justify each position from your data, not your taste. Then choose one CRAP principle and say exactly how you would apply it for the user type you added. **In pairs,** compare tables and agree on the top block.

Give evidence for your answer.

Stuck? Start with your tally: which need appeared most? Put that block at the top. Then ask of each block — highlight, unify, order, or group? Pick the CRAP principle that fixes the weakest spot for the user you added.

🌍 In a New Context

A school wants to redesign its cluttered results page so parents can find their child's grades quickly. Name the persona you would define, one wireframe decision, and one CRAP principle you would apply, and say what each one improves.

❓ Lesson Question — Answered

The lesson asked how designers plan a website around its users and what makes a page clear. Design begins from the user, not the creator. A persona pictures a typical user as a concrete person — age, occupation, purpose, and challenges — so the team can focus on what that user needs. A wireframe then fixes the placement and priority of information before any color or decoration, which keeps discussion on structure and helps stakeholders agree. To make the page clear, four principles are applied: contrast to highlight priority, repetition for consistency, alignment for order, and proximity to group related items. All of this is guided by user-centered design (UCD), which starts from what the user needs and aims to improve usability and accessibility. Good design, therefore, is planned from the user inward — who they are, what goes where, and how to make it clear.

KEY TERMS

usability — how easily users can use the site.

accessibility — usable by a wide range of people.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

When designing a website, it is important to view it from the user's perspective. First, create a (a), in which a typical user image is set up as a concrete person, and clarify "what this user needs." Next, create a (b), a blueprint for determining the placement and priority of information. In design, to improve viewability and clarity, the (c)—contrast, repetition, alignment, and proximity—should be kept in mind.

1. Fill in the blanks (a)–(c).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT included in the four principles of design (CRAP principles).

A Contrast
B Encryption
C Alignment
D Proximity

(2) From the following options (A–D), choose the one that most appropriately describes the concept of user-centered design.

A Prioritize the design that the developer finds easy to use.
B Place the highest importance on visual design.
C Begin the design from what the user needs.
D Use as many of the latest technologies as possible.

⭐ REFLECT & CHALLENGE

Reflect: On a website you find confusing, which CRAP principle seems most broken, and why? **Challenge:** Your clinic can fund only one accessibility improvement this year. For the low-vision or no-Internet user you identified, choose one CRAP principle to apply — and say what you would give up (a decoration or feature) to make room for it, and why that trade-off serves the user better.

⭐ Key takeaway

Remember:
Persona (who) → wireframe (where) → CRAP (make it clear) → UCD (decide for the user). Design from the user inward.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Lesson 4-3

Methods for Evaluating Websites

Learning Objectives

- Distinguish qualitative from quantitative evaluation, and explain the major indicators used in web analytics.
- Explain how heuristic evaluation works, and why combining both approaches gives a clearer picture of a website.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

How do you know whether a website works well for its users? You can watch real users try it and ask what confuses them — that is qualitative evaluation. Or you can count how many people visit, how long they stay, and how many buy — that is quantitative evaluation. Numbers tell you how much is happening; users' voices tell you why. The best teams use both, together with expert checks, to see a site clearly and improve it.

? Guiding question: How can a website be evaluated, and why is it useful to combine numbers with users' voices?

Key Fact:
Websites are evaluated qualitatively (why — usability tests, interviews, heuristic checks) and quantitatively (how much — web analytics: PV, bounce rate, CVR); combining both gives the clearest picture.

Key Concepts:

- qualitative evaluation
- quantitative evaluation
- web analytics
- PV / bounce rate / CVR
- heuristic evaluation

KEY TERMS
qualitative evaluation — the "why" from users' voices and behavior.
quantitative evaluation — the "how much" from numerical data.

Qualitative watches users; quantitative reads numbers.

Point!

1. Qualitative and Quantitative Evaluation

(1) Methods for evaluation generally fall into two main categories: **(qualitative evaluation)** and **(quantitative evaluation)**. These evaluation methods are used in a wide range of applications, and the same concepts apply to website evaluation.

Evaluation method	Description (purpose)	Specific techniques
Qualitative evaluation	Explore in depth the "reasons" behind users' psychology and behavior, and evaluate qualitative aspects that cannot be expressed in numbers.	• Usability testing: observe actual operations to identify points that are hard to use. • Interviews: hear users' direct voices to understand their needs and dissatisfactions.
Quantitative evaluation	Use numerical data such as access counts and time, and evaluate the current state objectively and statistically.	• Web analytics: analyze metrics such as PV count and time spent. • A/B testing: compare two options to verify which leads to better results.

(2) Qualitative evaluation is suited to exploring the cause of "why is it hard to use," while quantitative evaluation is suited to grasping the overall trend of "how much is it being used." By combining both, more accurate evaluation is possible.

2. Major Indicators in Web Analytics

(1) **(Web analytics)**…a technique for analyzing how a website is accessed using numerical values; it is the central method of quantitative evaluation for websites. The following indicators are commonly used.

PAUSE & THINK

Would counting visits tell you *why* people leave a page? Which method would?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Indicator	Meaning	Description
(PV (Page View))	The total number of times a page has been viewed.	A basic indicator that measures how widely the site is seen and how much exposure it gets.
(Bounce rate)	The proportion of visits that ended after viewing only the first page.	An indicator of whether the page content matched the user's expectations and whether further actions were prompted.
(Conversion rate (CVR))	The proportion of visits that achieved a target action (purchase, registration, etc.).	A very important indicator that measures how much the site contributes to business results (profits).

3. Heuristic Evaluation

(1) **(Heuristic evaluation)**…a method in which experts identify problems in a website's UI (user interface) based on rules of thumb (heuristics). It is a form of qualitative evaluation, used along with usability testing and other methods to investigate the causes of usability problems.

(2) Major evaluation perspectives

① Consistency…whether the design and operations are unified across the site.

② Error prevention…whether the design makes it hard for users to make mistakes.

③ Visibility of system status…whether system status is clearly indicated to users (e.g., processing indicators, input error messages).

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Qualitative evaluation is a method that uses numerical data to evaluate quantitatively.

B PV (Page View) is an indicator showing the number of times a page has been viewed.

C The bounce rate is an indicator showing the proportion of visits that achieved a target action.

D Heuristic evaluation is a method in which experts identify UI problems based on rules of thumb.

(2) Match each description (a–c) with the most appropriate indicator from the choices below (A–C).

a The number of times a page has been viewed

b The proportion of visits that ended after viewing only the first page

c The proportion of visits that achieved a target action

[Choices] **A** PV (Page View) **B** Bounce rate **C** Conversion rate

✓ Solution

Analytics tracks visits, bounce, and conversions.

KEY TERMS

PV — total page views.
bounce rate — the share of visits that end after only one page.
CVR — the share of visits that reach the target action.

Heuristic review checks a site against usability rules.

PAUSE & THINK

Match the question to the method: “how many?” vs “why?”

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(1)

A Qualitative evaluation observes user voices and behavior to evaluate qualitatively. Using numerical data is quantitative evaluation. Therefore, ×.

B Correct definition of PV. Therefore, ○.

C The bounce rate is the proportion of visits that ended after viewing only the first page. The proportion of target action achievement is the conversion rate. Therefore, ×.

D Correct definition of heuristic evaluation. Therefore, ○.

(2)

a: The number of page views is PV. **A**

b: The proportion who left after only the first page is the bounce rate. **B**

c: The proportion who achieved a target is the conversion rate. **C**

Try

1 Answer the following questions.

1. What is the term for the method that evaluates qualitatively by observing user voices and behavior?
2. What is the term for the method that evaluates quantitatively using numerical data?
3. What is the two-letter abbreviation for the indicator showing the number of times a page has been viewed?
4. What is the term for the method in which experts identify UI problems based on rules of thumb?

2 Answer the following questions.

(1) From the following options (A–D), choose all terms that are part of quantitative evaluation methods.

A Usability testing

B Web analytics

C Interviews

D A/B testing

EXAM-STYLE QUESTION

Explain why a website team should use both qualitative and quantitative evaluation rather than only one. **[6] Refer to:** what numbers show and what users' voices show.

PAUSE & THINK

Which of the four above rely on numbers, and which on people's words?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) From the following options (A–D), choose the one that most appropriately describes the conversion rate.

A The proportion of times a page has been viewed

B The proportion of visits that ended after viewing only the first page

C The proportion of visits that achieved a target action (product purchase, member registration, etc.)

D The proportion of viewed pages relative to the total number of pages on the site

Think as an Engineer — Investigate & Decide

A small online bookstore in your city wants to know whether its website really works for customers. You plan how to evaluate it.

☐ **Rely on the owner's opinion** ☐ **Use both quantitative and qualitative methods**

1 · Collect data. Ask five classmates who have used an online bookstore to visit any online shop and, for each, record one number you could measure (e.g., how many pages they viewed) and one thing they said was confusing. Put the results in a small table.

2 · Analyze the signals. Suppose the bookstore's PV is high but its conversion rate is very low. What might that mean — and which method (quantitative or qualitative) would tell you *why*? Then consider one customer the data may hide — for example, someone shopping only on a small phone screen, or a first-time visitor — and note one difficulty they might face that a raw PV count would never show.

3 · Decide. Recommend which two evaluation methods the bookstore should use first and give two reasons drawn from the table you built in step 1. **In pairs,** compare your choices and agree together on the single most useful indicator to track.

Give evidence for your answer.

Stuck? Start from the gap in the data: a high PV with a low CVR means people arrive but don't act — so the numbers only point to *where* to look. Pick the qualitative method (a short usability test or an interview) that answers the exact question your numbers raised.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

🌍 In a New Context

A school's new website gets many visits, but most people leave after the first page. Name one quantitative indicator that reveals this problem, one qualitative method to find its cause, and one likely fix.

❓ Lesson Question — Answered

The lesson asked how a website can be evaluated and why numbers and users' voices are combined. Evaluation falls into two kinds. Quantitative evaluation uses numerical data — through web analytics indicators such as PV (how many views), bounce rate (how many leave after the first page), and CVR (how many reach a target action) — to show how much a site is used. Qualitative evaluation explores the reasons behind users' behavior through usability testing, interviews, and heuristic evaluation, in which experts check the UI against rules of thumb. Numbers reveal what is happening but not why; users' voices and expert checks reveal the causes. Combining both — the "how much" and the "why" — gives the most accurate evaluation and the clearest path to improvement.

KEY TERMS

web analytics — numbers on how a site is accessed.
heuristic evaluation — experts check the UI by rules of thumb.

⭐ REFLECT & CHALLENGE

Reflect: If you could see only one metric for a site you use, would you choose PV, bounce rate, or CVR — and why?
Challenge: A page has a high bounce rate. Design a short qualitative test that would reveal the cause, and say what you would change first.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Methods for evaluating a website include (a) evaluation, which observes user voices and behavior, and (b) evaluation, which analyzes using numerical data. Typical indicators used in (b) evaluation are (c), which shows the number of times a page has been viewed; (d), which shows the proportion of visits that ended after viewing only the first page; and (e), which shows the proportion of visits that achieved a target action.

1. Fill in the blanks (a)–(e).

3 Answer the following questions.

(1) For each of evaluation methods (a–c), indicate which it falls under: **A** Qualitative evaluation, **B** Quantitative evaluation.

a Have users actually explore the website and observe its usability

b Tally the access counts and page views of the website

c Interview users to hear their impressions and improvement requests

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) From the following options (A–D), choose the one that most appropriately describes heuristic evaluation.

A A method in which users explore the site and usability is observed

B A method in which numerical data is tallied and site usage is analyzed

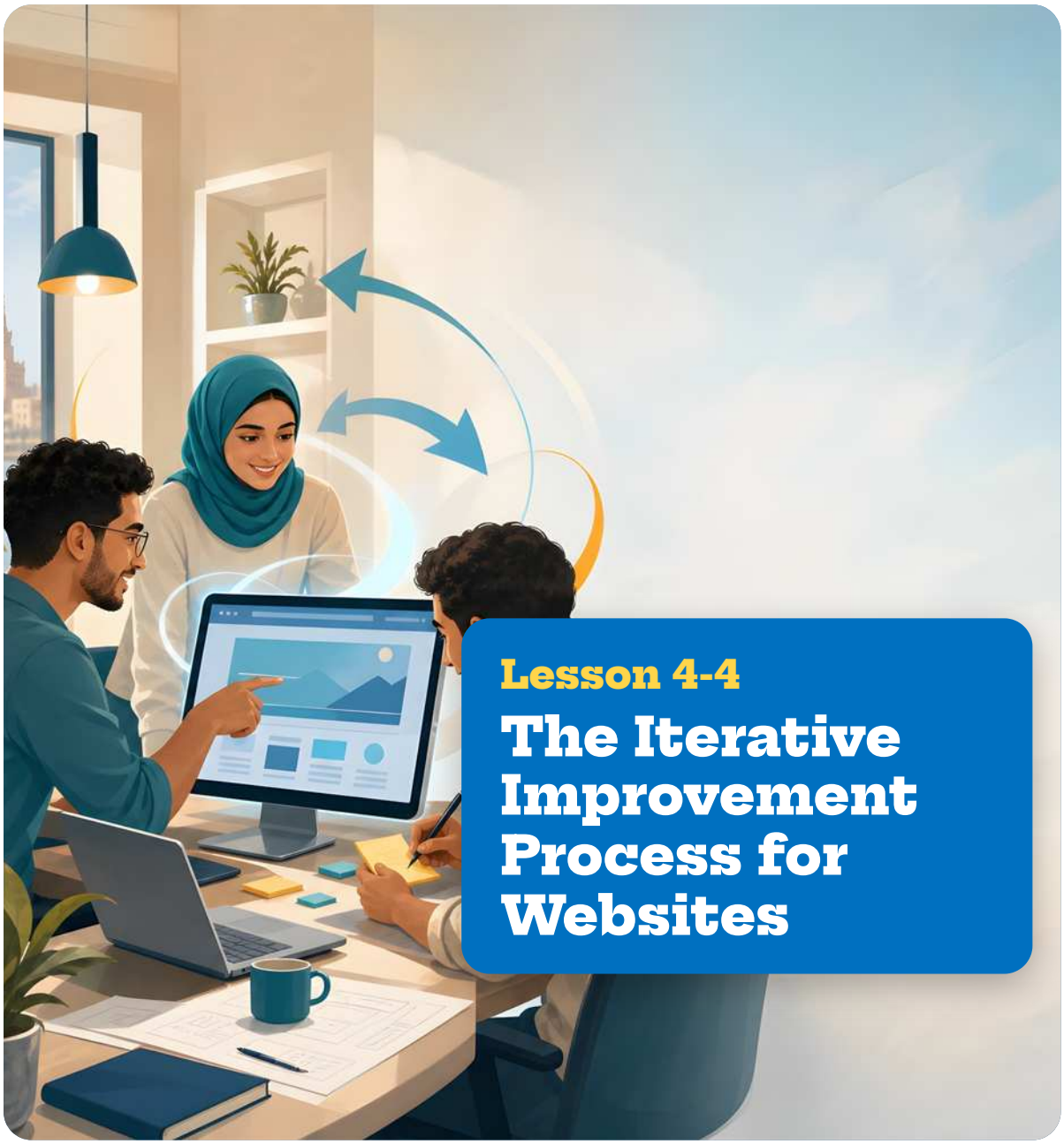
C A method in which experts identify UI problems based on rules of thumb

D A method in which two patterns are prepared and compared with data to see which is more effective

Key takeaway

Remember:
Quantitative (PV, bounce rate, CVR) shows how much; qualitative (usability tests, interviews, heuristic checks) shows why. Combine both to evaluate a site clearly.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 4-4

The Iterative Improvement Process for Websites

Learning Objectives

- Explain the PDCA cycle and the steps of design thinking as ways of improving a website.
- Describe how A/B testing compares two options using real data, and explain why improvement is repeated in cycles.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

A website is never truly finished. The first version is only a best guess, and a team learns what truly works only after real people use it. That is why good teams improve in loops: they plan a change, make it, measure the result, and act on what they learn — the PDCA cycle. They also design from the user's side (design thinking) and test real choices with data (A/B testing). Small, repeated improvements are better than waiting for one perfect launch.

? Guiding question: Why is improving a website a repeated loop rather than a one-time job, and how do teams decide what to change?

Key Fact:
Websites improve through repeated cycles: PDCA (Plan, Do, Check, Act) drives continuous improvement; design thinking finds problems from the user's side; A/B testing compares two options with real data.

Key Concepts:

- PDCA cycle
- Plan / Do / Check / Act
- design thinking
- A/B testing

KEY TERMS
PDCA cycle — Plan→Do→Check→Act, repeated.
continuous improvement — getting better a little at a time.

📌 **Point!**

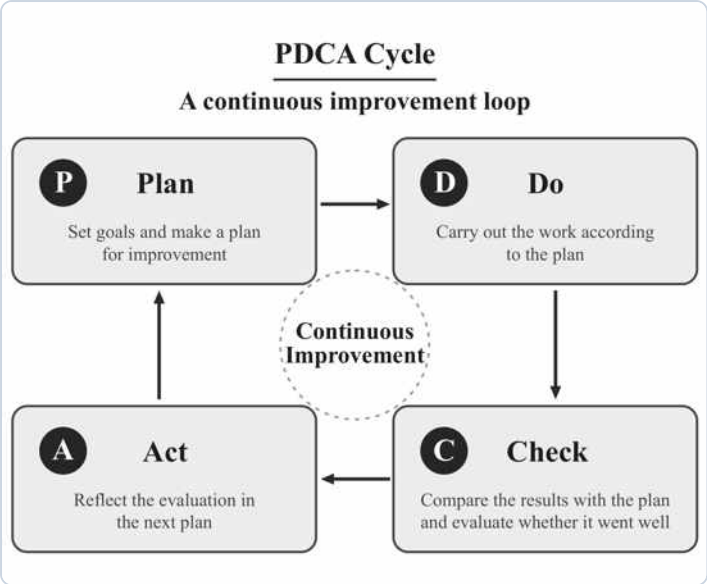
1. The PDCA Cycle

(1) **(PDCA cycle)**…a method that continuously advances improvement by repeating the following four steps. It is used widely in fields such as business improvement and quality management, not just websites.

Step	Description	Example in website improvement
Plan	Set goals and create a plan for improvement	Set the goal “reduce bounce rate by 10%” and plan a redesign of the homepage
Do	Execute according to the plan	Change the design of the homepage
Check	Measure results and evaluate the level of goal achievement	Measure the bounce rate after the change and evaluate whether the goal (10% reduction) was achieved
Act	Based on the evaluation, consider the next improvement measures	If the effect is sufficient, continue; if not, consider next steps such as revising headings or image placement

Each Plan-Do-Check-Act cycle improves the site.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



The PDCA cycle as a continuous improvement loop

(2) The important point of the PDCA cycle is not to aim for completion in one round, but to gradually improve through repetition.

2. Design Thinking

(1) **(Design thinking)**…a way of thinking that leads to discovering problems and generating solutions from the perspective of the user. It proceeds through the following five steps.

Step	Description
① **Empathize**	Put yourself in the user's position and understand their challenges and needs
② **Define**	Based on the information gleaned from empathy, clarify the problem to be solved
③ **Ideate**	Generate many ideas to solve the problem
④ **Prototype**	Quickly create a prototype based on the ideas
⑤ **Test**	Have users try the prototype and obtain feedback from them

(2) Based on test results, one may go back to the empathize or define stage. Improvement proceeds back and forth, not one-way.

3. A/B Testing

(1) **(A/B testing)**…a method in which two patterns (option A and option B) are prepared, each is shown to actual users, and data is examined to determine which is more effective.

e.g., Compare option A with a red purchase button and option B with a blue purchase button to see which gets clicked more.

(2) Key points of A/B testing

PAUSE & THINK

Which PDCA step tells you whether your change actually worked?

Design thinking: ideate, prototype, test, repeat.

A/B testing compares two versions with real users.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Change only one element at a time (if multiple elements are changed, it is unclear which change produced the effect).

Make decisions only after collecting a sufficient amount of data.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A The PDCA cycle is a method that repeats the four steps Plan→Do→Check→Act.

B Design thinking is a way of thinking that solves problems from the developer's perspective.

C In A/B testing, it is important to change only one element at a time.

D The PDCA cycle is a method that is complete after one execution.

(2) For each of a–e, indicate which of the five steps of design thinking (A–E) it corresponds to.

a Understand the user's challenges and needs.

b Clarify the problem to be solved.

c Generate many ideas.

d Quickly create a prototype.

e Have users try the prototype.

A Empathize **B Define** **C Ideate** **D Prototype** **E Test**

✓ Solution

(1)

A Correct definition of the PDCA cycle. Therefore, ○.

B Design thinking is a way of thinking from the user's perspective. Therefore, ×.

C It is important to limit the changed element to one in A/B testing. Therefore, ○.

D The PDCA cycle is meaningful when repeated. Therefore, ×.

(2)

a: Understanding the user is empathize. **A**

b: Clarifying the problem is define. **B**

c: Generating ideas is ideate. **C**

d: Creating a prototype is prototype. **D**

e: Having users try is test. **E**

KEY TERMS

design thinking — solve from the user's view in five steps.

A/B testing — compare two options with real data.

💡 PAUSE & THINK

In A/B testing, why change only one element at a time?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the term for the improvement method that repeats the four steps Plan→Do→Check→Act?
2. What is the term for the way of thinking for discovering problems and generating solutions from the perspective of the user?
3. What is the term for the method in which two patterns are prepared and compared with data to see which is more effective?
4. Name the five steps of design thinking in order.

2 Answer the following questions.

(1) From the following options (A–D), choose the one that corresponds to "Check" in the PDCA cycle.

A Set the goal of reducing the bounce rate by 10%.

B Change the design of the homepage.

C Measure the bounce rate after the change.

D Consider another improvement measure.

(2) From the following options (A–D), choose the one that is NOT an appropriate way of conducting A/B testing.

A Prepare two patterns and compare them.

B Change multiple elements at the same time.

C Make decisions only after collecting a sufficient amount of data.

D Show each pattern to actual users.

Think as an Engineer — Investigate & Decide

Your school club runs a web page for its yearly event, but few students sign up through it. You lead one round of improvement using the PDCA cycle.

☐ **Redesign everything at once** ☐ **Improve in one measured PDCA loop**

1 · Collect data (Check first). Decide one number you could measure now — for example, how many of ten classmates find the "sign up" link within five seconds. Record the result as your starting point.

2 · Plan one change. From the users you watched, name the single biggest obstacle (for a specific group, such as a student opening the page on a small phone screen) and plan one change to fix it. State the goal as a number (e.g., "8 of 10 find it in five seconds").

EXAM-STYLE QUESTION

Explain why a website team improves in repeated PDCA cycles instead of trying to build a perfect site in one go. **[6]**
Refer to: measuring results and acting on them.

PAUSE & THINK

Match each option above to a PDCA step: which is Plan, Do, Check, Act?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 · Do, Check, Act — and decide. Describe how you would make the change, measure again, and decide the next step. Choose one element to settle with an A/B test, and say why you change only that one element. **In pairs,** compare plans and agree on the single change to try first.

Give evidence for your answer.

Stuck? Start from your measured number (Check), pick the one obstacle it exposes, plan a single fix (Plan), then loop — do not change everything at once.

🌍 In a New Context

An online shop finds that its “buy” button is rarely clicked. Describe how you would use one A/B test (changing only one element) and one PDCA loop to improve it, and say which number would tell you it worked.

KEY TERMS

prototype — a quick, testable draft.

iterate — repeat the loop to improve.

❓ Lesson Question — Answered

The lesson asked why improving a website is a repeated loop and how teams decide what to change. A first version is only a guess, so teams improve continuously rather than once. The PDCA cycle repeats four steps — Plan (set a goal and plan a change), Do (make it), Check (measure the result), and Act (decide the next step) — and its point is not completion in one round but gradual improvement through repetition. To decide what to change, design thinking works from the user's side through five steps (Empathize, Define, Ideate, Prototype, Test), moving back and forth rather than one-way, while A/B testing compares two options with real data, changing only one element at a time and deciding only after enough data. Together they make improvement small, measured, and repeatable.

Exercise

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

As a method for improving websites, there is the (a) cycle. It repeats four steps: (b), in which goals are set; (c), in which the plan is executed; (d), in which results are measured; and (e), in which improvement measures are considered. Also, (f), in which two patterns are prepared and compared with data to see which is more effective, is a useful method.

1. Fill in the blanks (a)–(f).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes design thinking.

A A method that uses numerical data to analyze website usage.

B A way of thinking that leads to discovering problems and generating solutions from the perspective of the user.

C A method in which experts identify UI problems based on rules of thumb.

D A method of creating a blueprint that roughly shows the layout of a web page.

(2) From the following options (A–D), choose the one that most appropriately describes the concept of the website improvement process.

A If a perfect website is built initially, no improvement is needed.

B It is enough to improve only the visual design.

C It is important to gradually improve by repeating evaluation and improvement.

D All improvements should be done with A/B testing alone.

⭐ REFLECT & CHALLENGE

Reflect: Think of a website you use often. Which one small change do you think would help most, and how would you measure whether it worked?
Challenge: You may run only one A/B test this month on the event page. Choose the single element to test, predict which option will win and why, and say what number would prove you right or wrong.

⭐ Key takeaway

Remember:
Improve in loops: PDCA (Plan, Do, Check, Act) for continuous improvement, design thinking (Empathize, Define, Ideate, Prototype, Test) from the user's side, and A/B testing to settle one change with data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Editorial Team

Project Lead & Editorial Director

Masaki Iisaka

Editors & Writers

Biichi Nakajima
Hiromi Fujii

Reviewers

Prof. Mohamed Sayed Farag.

Prof. Eman Serag El-Din Bakr.

Dr. Abeer Hamed Ahmed.

Dr. Taher Abdel Hamid El-Adly

Dr. Manal Ziada Abdel Fadil.

The Central Administration for Curriculum Development

General Supervision

Dr. Gebriel Anwer Hemieda

Head of the Central Administration for Curriculum Development

All Rights Reserved ©Ministry of Education and Technical Education - 2026 / 2027

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means-electronic, photocopying, recording, or otherwise-without prior written permission of the Ministry of Education and Technical Education.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

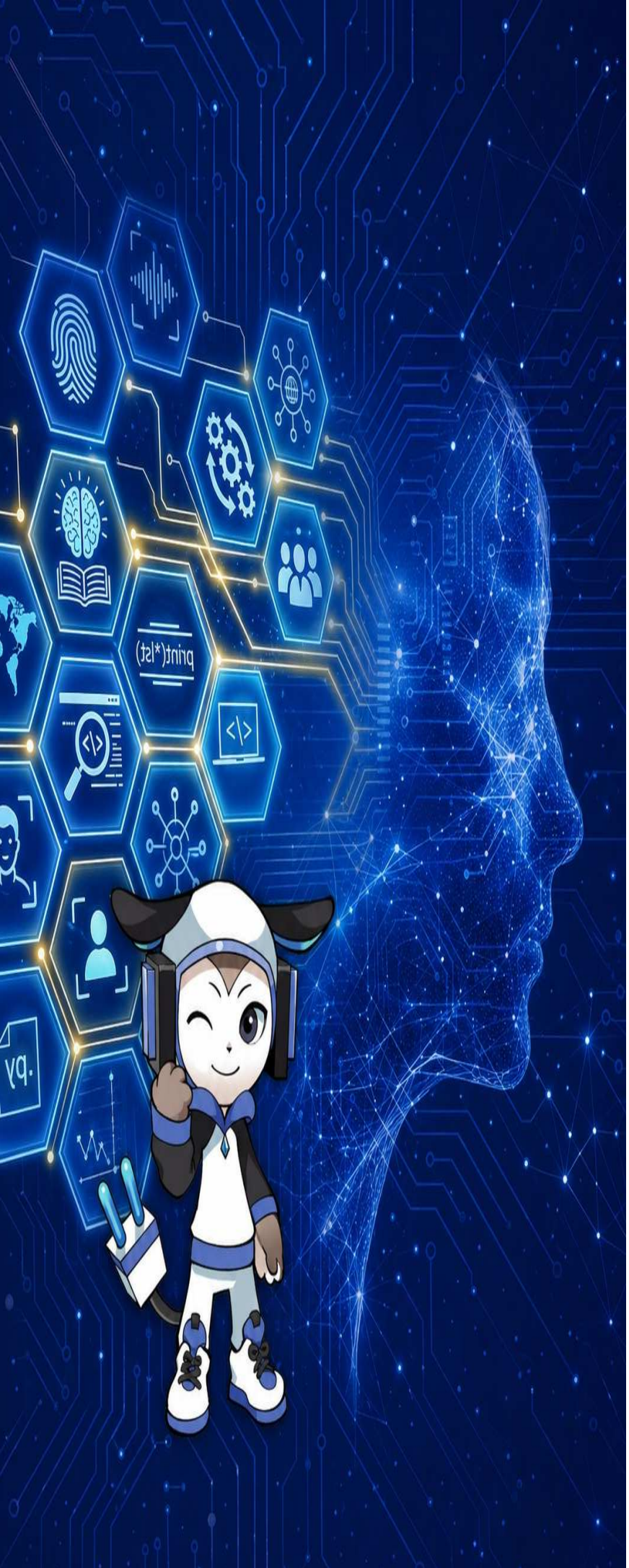
EDUCATION SPRINGS NEW LIFE

SPRIX

CACD

class	name

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Programming and Artificial Intelligence

An Introduction to
Information and Communication Technology

Egyptian Baccalaureate

2nd Year

Part Two

11

SPRIX

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Programming and Artificial Intelligence

Egyptian Baccalaureate

2nd Year

The International Baccalaureate™

has reviewed the pedagogical and assessment approaches underpinning this textbook.

© 2026 Ministry of Education and Technical Education

All rights reserved.

Copyrights and accountability for the content of this textbook remain with the Ministry of Education and Technical Education.

No part of this publication may be reproduced, stored in a retrieval system,

or transmitted in any form or by any means - electronic, mechanical, photocopying, recording,

or otherwise - without the prior written permission of the copyright holders.

Ministry of Education and Technical Education

New Administrative Capital

Cairo, Egypt

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Introduction

In light of Egypt's Vision 2030 and the state's ongoing efforts to develop education, the Ministry of Education and Technical Education places great emphasis on updating school curriculums — particularly in information and communication technology — due to its vital role in fostering computational thinking, problem-solving abilities, and responsible digital citizenship, with a focus on building a deep understanding of fundamental concepts rather than rote memorization. The content is organized in a gradual and systematic manner that enables students to progress smoothly from simple to complex concepts, linking every concept to real-life situations, and relying on interactive methods that include teamwork and discussion, making learning more enjoyable.

In this context, the Ministry has been keen to benefit from successful international experiences in teaching information technology, data analysis and artificial intelligence, given the speed at which these fields are reshaping work, communication and society, and the need for curriculums that keep pace with them while remaining clear and accessible to every student.

From this concern, this book offers an integrated introduction to information and communication technology. Part 1 addresses information technology and society, cybersecurity, web applications, and web and media design; Part 2 addresses data collection and cleaning, analysis and communication, and machine learning and artificial intelligence — so that the student meets data and AI as tools to be understood, questioned and used with judgement.

This book represents a step within a comprehensive plan to develop information and communication technology curriculums across all grade levels, with the aim of building an integrated educational system that combines the best international practices with Egypt's authentic identity.

The Ministry of Education and Technical Education firmly believes that developing education is a true investment in the nation's future, and that updating curriculums is not an end in itself, but rather a means to build an aware, creative generation, capable of keeping pace with scientific and technological changes in today's world.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Contents

5 Data Collection and Cleaning 4

5-1 Methods of Data Collection 4

5-2 Data Cleaning and Transformation 12

5-3 Open Data and APIs 21

6 Analysis and Communication 29

6-1 Statistical Inference 29

6-2 Use and Evaluation of Regression Analysis 39

6-3 Data Visualization and Communication 48

7 Machine Learning and Artificial Intelligence 57

7-1 The Basics of Machine Learning 57

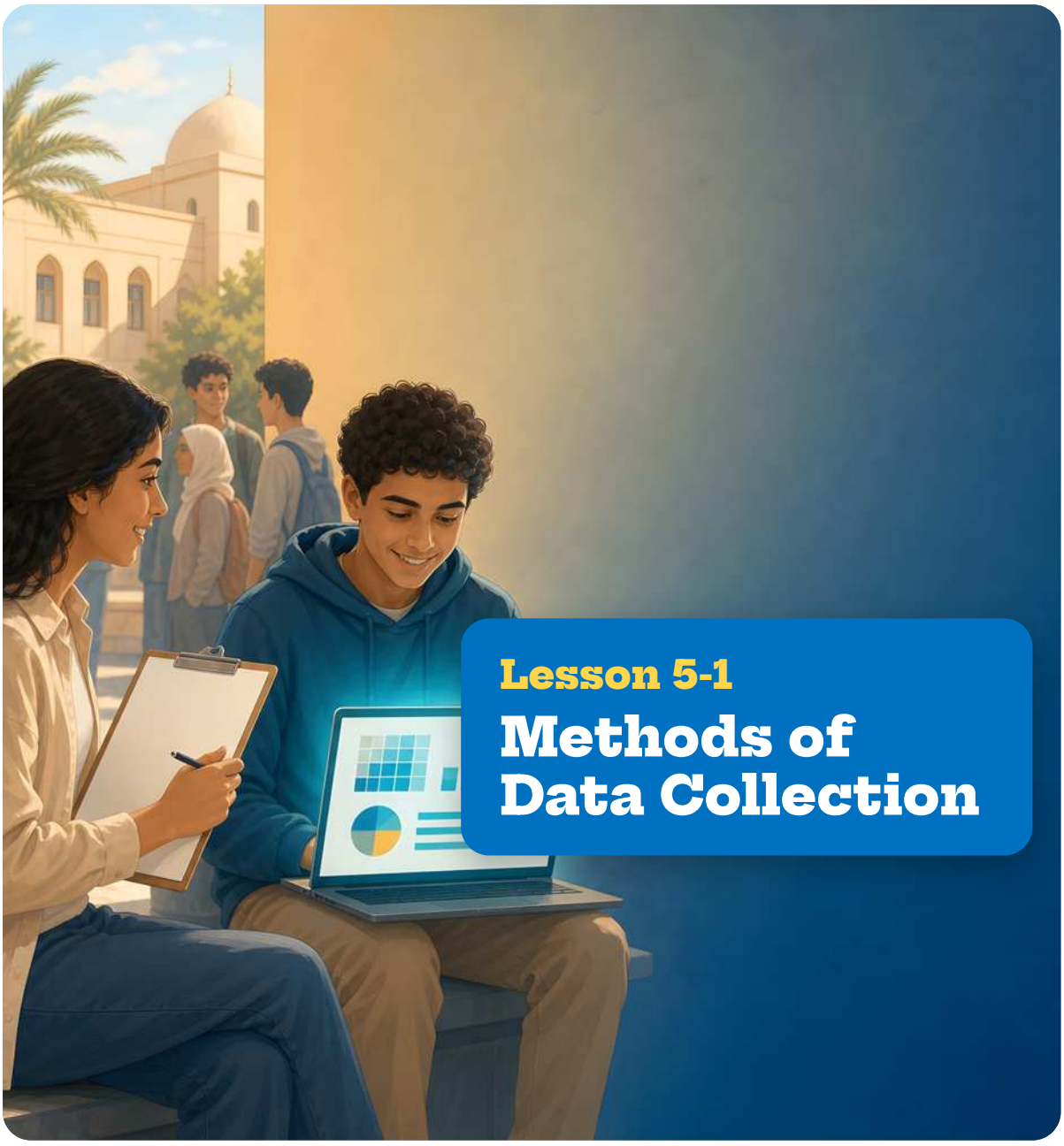
7-2 Neural Networks and Deep Learning 68

7-3 Large Language Models (LLM) and Generative AI 77

Answers 86

Glossary and Index 88

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Learning Objectives

- Distinguish primary from secondary data and explain when each is appropriate.
- Explain how sampling works, and describe how sampling bias can lead to confident but wrong conclusions.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Before you can trust what data tells you, you have to trust how it was gathered. Good analysis moves through three stages — collection, organization, then analysis — and if the wrong data is collected, no careful work later can rescue it. Some data you gather yourself for your exact purpose (primary data); some already exists, collected by others (secondary data). And when a whole group is too large to survey, you study a sample — but only a fair sample lets you speak for the whole. A biased sample gives a confident answer that is simply wrong.

? Guiding question: How do you collect data you can trust, and how can a small sample fairly represent a whole population?

📌 Point!

1. The Data Collection Stage

(1) Data analysis progresses through the flow of "collection → organization → analysis," after the purpose has been clarified. In **(data collection)**, the first stage, it is necessary to plan what data will be gathered and how, in line with the purpose of the analysis.

(2) If data that does not match the purpose is collected, it is impossible to draw appropriate conclusions, no matter how carefully later organization and analysis are carried out. For this reason, data collection is an important stage that determines the quality of the entire analysis.

2. Primary Data and Secondary Data

(1) **(Primary data)**…data that is newly collected to fit a specific purpose. *e.g.,* surveys, interviews, observations, experiments. Strength: the items collected can be designed to match the purpose, providing excellent originality and reliability. Constraint: collection takes time, effort, and cost.

(2) **(Secondary data)**…data that has already been collected and made public by others, and is used in analysis. *e.g.,* open datasets, existing statistics, industry reports. Strength: large amounts of data can be obtained in a short time. Constraints: it may not perfectly match one's purpose; bias or omissions may have arisen during editing or summarizing.

(3) Primary data, since it can be collected to match one's own purpose, has excellent reliability and originality, making it a high-value type of data that can serve as the core of analysis. On the other hand, the value of secondary data lies in being able to easily make use of large-scale data and existing statistics that cannot be obtained from primary data alone. In actual analysis, the two are appropriately combined depending on the purpose and constraints.

Key Fact:
Good analysis starts with good collection: primary data is gathered fresh for a purpose; secondary data already exists; sampling studies part of a population to estimate the whole — and only an unbiased sample supports valid conclusions.

Key Concepts:

- data collection
- primary / secondary data
- population & sample
- sampling (simple random, stratified)
- sampling bias

Sound collection is the base of trustworthy data.

KEY TERMS
data collection — the first stage: plan what to gather and how.
primary data — collected fresh for a purpose.
secondary data — already collected by others.

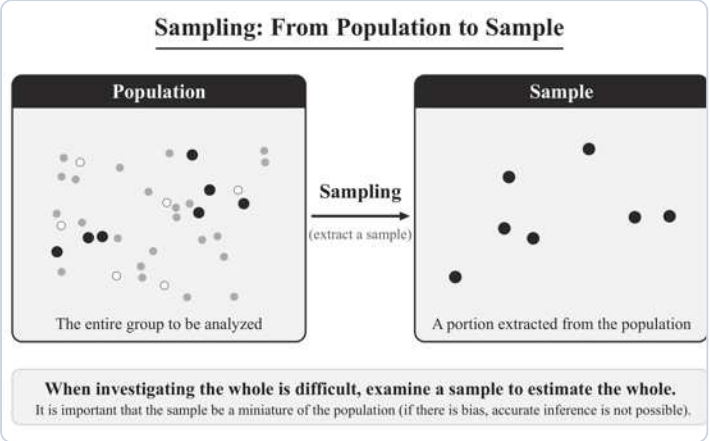
Primary data is gathered fresh; secondary is reused.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3. Sampling

(1) The entire group that is the target of analysis is called the **(population)**, and the part of the population that is actually surveyed is called the **(sample)**. Surveying the entire population is called a **(complete survey)**, also called a census, and surveying a chosen sample is called a **(sample survey)**.

(2) **(Sampling)**…selecting a sample from the population. Depending on the sampling method, the reliability of the conclusions obtained varies greatly.


Sampling: From Population to Sample
Population
The entire group to be analyzed
Sampling
(extract a sample)
Sample
A portion extracted from the population
When investigating the whole is difficult, examine a sample to estimate the whole.
It is important that the sample be a miniature of the population (if there is bias, accurate inference is not possible).

Sampling: from population to sample

(3) A typical sampling technique is **(random sampling)**, as described below. Random sampling makes it likely that the sample represents the population, allowing the analysis results to be applied to the population as a whole.

Method	Description
(Simple random sampling)	Choose a sample from the population purely by chance, without bias
(Stratified sampling)	Divide the population into strata such as gender or age group, and randomly choose from each stratum

(4) Simple random sampling is the most basic method and can be used even when little is known about the population. Stratified sampling can reflect the composition of the population (such as the ratios of gender and age group) in the sample, making it easier to reduce bias when differences between strata affect the analysis.

4. Bias in Surveys

(1) **(Sampling bias)**…this occurs when the method of selecting a sample is biased and does not correctly reflect the actual state of the population. When sampling bias is present, the conclusions obtained cannot be applied to the population as a whole.

(2) Typical examples of bias. Self-selection and time/location bias concern who is surveyed (selection bias); response bias concerns how those surveyed answer (measurement error).

PAUSE & THINK

A government census you download — is that primary or secondary data for you?

A fair sample mirrors the whole; a biased one misleads.

KEY TERMS

population — the whole group analyzed.
sample — the part actually surveyed.
sampling bias — an unfair sample that misrepresents the whole.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Type of bias	Description	Example
(Self-selection bias)	Bias caused by only those who want to respond actually responding	In a web survey, only people with high interest respond
(Response bias)	Not a selection problem: people already surveyed give answers that look desirable rather than accurate	Reporting a “study time” longer than it really was
(Time/location bias)	Bias caused by the timing or location of the survey being limited	Surveying in front of a station on weekday afternoons, missing office workers and students

(3) It is difficult to eliminate sampling bias completely, but bias can be reduced by carefully designing the survey: how subjects are chosen, how questions are worded, and how the survey is conducted.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Primary data refers to data that has already been collected and made public by others.

B Selecting a sample from a population is called sampling.

C Stratified sampling is a method in which the population is divided into strata, and a sample is randomly chosen from each stratum.

D When sampling bias is present, the conclusions obtained may not be applicable to the population as a whole.

(2) Match each description (a, b) of a sampling method with the most appropriate option from the choices below (A, B).

a Choose a sample from the population by chance, without bias.

b Divide the population into strata such as gender or age group, and randomly choose from each stratum.

[Choices] **A** Stratified sampling **B** Simple random sampling

✓ Solution

(1)

A Primary data is data that is newly collected to fit a specific purpose. Data already collected and made public by others is secondary data. Therefore, ×.

B Correct definition of sampling. Therefore, ○.

C Correct definition of stratified sampling. Therefore, ○.

D Correct description of the effect of sampling bias. Therefore, ○.

PAUSE & THINK

A web poll anyone can enter — whose voices are missing, and which bias is that?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2)

a: Choosing without bias by chance is simple random sampling. **B**

b: Dividing into strata and choosing randomly from each is stratified sampling. **A**

Try

1 Answer the following questions.

1. What is the term for data that is newly collected to fit a specific purpose?
2. What is the term for data that has already been collected and made public by others, and is used in analysis?
3. What is the term for selecting a sample from a population?
4. What is the term for when the method of selecting a sample is biased and does not correctly reflect the actual state of the population?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is most appropriate as a method of collecting primary data.

A Use statistical data published by a government agency.

B Create a survey to fit one's own purpose and collect responses.

C Cite figures from a report issued by an industry association.

D Use survey data from a study conducted earlier by someone else as is.

(2) For each survey situation (a, b), choose the most appropriate sampling method from the choices below (A, B).

a When investigating study time across all students in a school, you want to choose the sample taking into account differences in tendencies between grades.

b You want to grasp the opinions of all eligible voters in a city without bias.

[Choices] **A** Simple random sampling **B** Stratified sampling

EXAM-STYLE QUESTION

Explain why a biased sample can make an analysis worthless even when the calculations are correct. **[6] Refer to:** the sample as a miniature of the population.

PAUSE & THINK

Which option gathers new data *you* designed — and which just reuses others' data?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that is most appropriate as an example in which sampling bias has occurred.

A Surveying 100 students randomly chosen from all students in the school by drawing lots.

B Dividing all residents into age-group strata and surveying a randomly chosen sample from each stratum.

C Approaching passers-by in front of the station on weekday afternoons and reporting the results as “the opinions of all citizens.”

D Designing the wording of a survey question carefully so that respondents can answer truthfully.

Think as an Engineer — Investigate & Decide

Your school wants to know how long students typically spend commuting to school each day, but you cannot survey everyone. You design how to choose and question a sample.

☐ **Ask whoever is easiest to reach**

☐ **Plan a fair, representative sample**

1 · Collect data (plan the sample). Decide how many students you would ask and how you would pick them so the sample fairly represents the whole school. As a quick pilot, tally the commute of 8–10 classmates. State the population and the sample clearly.

2 · Analyze the risk of bias. Name one group your method might silently drop — for example, students who were absent that day, or students who never open an online survey link. Explain how missing that group would skew the estimated commute, and say which bias it is closest to (self-selection, response, or time/location).

3 · Decide. Choose simple random or stratified sampling for this survey and justify the choice from the risk you found in step 2 — think about whether commute really differs between identifiable subgroups here. **In pairs,** compare methods and agree on the single change that would most reduce bias.

Give evidence for your answer.

Stuck? Start from who might be left out. Ask yourself: could one part of the school answer very differently from the rest — a whole group, or students who never see your survey? If a subgroup might differ systematically, the real question is how to give every subgroup a fair chance *before* you pick your method. Work out which subgroups matter here, then let that guide your choice.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

🌍 In a New Context

A shopping mall surveys only visitors on Saturday afternoon and reports the results as “what all our customers think.” Name the bias, explain who is missed, and describe one change that would make the sample fairer.

❓ Lesson Question — Answered

The lesson asked how to collect data you can trust and how a small sample can represent a whole population. Trustworthy analysis begins with collection: after the purpose is clear, you choose primary data (gathered fresh for that purpose, with high reliability but more cost) or secondary data (already collected by others, fast to obtain but perhaps mismatched or biased), often combining both. When the population — the whole group analyzed — is too large, you study a sample through sampling, aiming for a miniature of the population. Random sampling makes that likely: simple random sampling picks purely by chance, while stratified sampling splits the population into strata and picks from each to mirror its make-up. The danger is sampling bias, where a skewed selection cannot speak for the whole; it is reduced by careful design of who is chosen, how questions are worded, and how the survey is run.

KEY TERMS

simple random sampling — picked purely by chance.
stratified sampling — split into strata, then pick from each.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

In data analysis, it is important to plan what data to collect according to the purpose. The data to be collected can be classified into two types: (a), which is newly collected, and (b), which has already been collected and made public by others. Also, when investigating the entire population is difficult, (c) is performed to select a part of the population. At this time, if bias arises in the way of selection, (d) occurs, and the analysis results may not be applicable to the population as a whole.

1. Fill in the blanks (a)–(d).

⭐ REFLECT & CHALLENGE

Reflect: Think of an online poll you have seen. Whose voices were probably missing from it, and why? **Challenge:** You may survey only 40 of your school's 800 students about a new schedule. Describe exactly how you would choose those 40 to avoid bias, and name the one group you would make sure is represented.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 Answer the following questions.

(1) From the following options (A–D), choose the one that is most appropriate as an example of secondary data.

A Survey responses collected from students in one's own class.

B Population statistics published by the government.

C Traffic on a school commute route observed and recorded by oneself.

D Records of an interview designed and conducted by oneself.

(2) From the following options (A–D), choose the one that most appropriately describes sampling methods.

A Simple random sampling is a method in which the surveyor preferentially chooses subjects that are easy to gather.

B Stratified sampling is a method in which the population is divided into strata, and a sample is randomly chosen from each stratum.

C Simple random sampling is a method in which the population is divided into strata, and a representative is chosen from each stratum.

D A complete survey is a method in which only part of the population is surveyed.

(3) From the following options (A–D), choose the one that is NOT appropriate as a way to reduce sampling bias.

A Choose subjects without bias from the entire population.

B Limit the timing and location of the survey, and gather responses only from people in specific conditions.

C Choose the sample taking into account the strata of the population (age group, gender, etc.).

D Design the wording of survey questions so that respondents can answer truthfully.

⭐ Key takeaway

Remember:
Collect the right data (primary vs secondary), study a fair sample of the population, use random sampling (simple or stratified) so the sample mirrors the whole, and guard against sampling bias.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Learning Objectives

- Explain how missing values, outliers, and duplicate data are handled when cleaning a dataset.
- Describe data type conversion, normalization and standardization, and the reshaping of data to fit a purpose.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

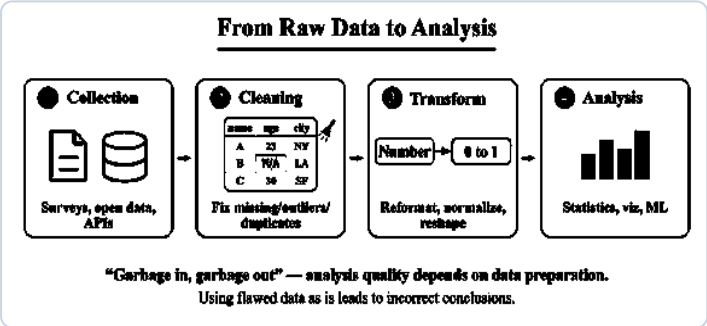
Real data arrives messy: blank cells, a value ten times too large, the same row entered twice, dates written three different ways. Before any analysis can be trusted, the data must be cleaned and transformed to fit the question. There is a rule in computing — "garbage in, garbage out" — which means careful calculation on flawed data still gives a flawed answer. This lesson is about turning raw data into data you can actually rely on.

? Guiding question: Why must data be cleaned and transformed before analysis, and what choices does each fix involve?

Point!

1. The Data Cleaning and Transformation Stages

(1) Collected data often cannot be used for analysis as is. Adjusting issues such as missing values, outliers, and duplicate data, and arranging the data into a format that fits the purpose of the analysis, is called **(data cleaning)** and **(data transformation)**.



From raw data to analysis: collect, clean, transform, analyze

(2) If issues are left unaddressed during analysis, incorrect conclusions may be drawn. For this reason, cleaning and transformation are important stages that support the accuracy and reliability of the analysis.

2. Handling Missing Values

(1) Using data containing missing values (values where data is absent) directly in analysis can prevent calculations from being performed correctly or cause biased results. For missing values, one of the following three responses is chosen.

Key Fact:
Raw data is rarely ready to use. Cleaning fixes missing values, outliers, and duplicates; transformation converts types, rescales, and reshapes data to fit the purpose. "Garbage in, garbage out" — the quality of any analysis depends on this preparation.

Key Concepts:

- data cleaning / transformation
- missing values, outliers, duplicates
- data type conversion
- normalization / standardization
- reshaping (pivot, join, aggregation)

KEY TERMS
data cleaning — fixing missing values, outliers, duplicates.
data transformation — reshaping/rescaling data to fit the purpose.

PAUSE & THINK
A blank cell in a survey — delete the row, fill it in, or flag it? What decides?

Fix blanks and judge outliers before analysis.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Response	Description	When to apply
Deletion	Remove rows or columns containing missing values from the data	When there are few missing values and the impact on the whole is small
Imputation	Fill in the missing parts with values such as the mean, median, or surrounding values	When missing values are scattered without bias, and replacing them with representative values from other data does not significantly change the overall trend
Flagging	Leave a marker indicating that the value was missing, and analyze the data as is	When the missingness itself has meaning (e.g., the fact that someone refused to answer)

(2) Which response to choose is determined by the proportion and pattern of missing values, and the purpose of the analysis. For example, imputing the mean for an item where missing values account for a large proportion of all data may produce values that are far from reality. There is no single correct response; judgment according to the purpose is required.

3. Handling Outliers and Duplicate Data

(1) Outliers (values that deviate significantly from other data) are not necessarily errors. They may be **(abnormal values)** caused by input or measurement errors, or they may be values that reflect reality. *e.g.,* If the sales data of a certain store shows that one day has remarkably high sales, it may be an input error, or it may genuinely reflect a sale or event on that day.

KEY TERMS
missing value — a value that is absent.
outlier — a value far from the others.
abnormal value — an outlier caused by an error.

Remove only genuinely duplicated records.

(2) Handling outliers begins with identifying the cause.

① When the cause is an input or measurement error…correct or delete the value

② When it is a value that reflects reality…in principle, keep it and decide how to handle it based on the purpose of the analysis

(3) Duplicate data (a state in which the same content exists in multiple records) affects the results of aggregation and is, in principle, removed. However, it is important to confirm that the data is genuinely duplicated. Even if records appear identical, they may include data that should be retained as separate records. *e.g.,* If the same customer purchases the same product twice on the same day, the records are separate transactions; combining them into one would make the sales aggregation incorrect.

4. Data Type Conversion

(1) Collected data is sometimes difficult to handle in its original format. Arranging the format to make data usable for analysis is called **(data type conversion)**. Typical examples include the following.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Type of conversion	Description	Example
Conversion from string to numeric	Convert numbers stored as strings into numeric values that can be used in calculations	“123” → 123
Unification of date format	Align dates with different notations into a single format	“2026/4/1,” “2026-04-01,” “01 Apr 2026” → 2026-04-01
Unification of notation	Combine values with the same meaning into a single notation	A mix of “Male/Female” and “M/F” → standardized to “Male/Female”

(2) If formats are not aligned, data cannot be aggregated correctly. For example, if numbers are stored as strings, sums and means cannot be calculated; if date notations are not unified, data from the same day will be treated as belonging to different days. For this reason, data type conversion is a step that must be checked before analysis.

5. Normalization and Standardization

(1) When the units or ranges (scales) of data differ greatly between variables, variables with larger values may strongly influence the analysis results. To prevent this, processing is performed to align the scale of data. Typical methods are **(normalization)** and **(standardization)**.

Method	Description	Main purpose
Normalization	Convert data to a range from 0 to 1	Align data with different units into a comparable form
Standardization	Convert data to a form with mean 0 and standard deviation 1	Position values relative to the spread of the data

(2) For example, when analyzing variables with different units and scales together, such as height (cm) and weight (kg), normalization can align both into the range from 0 to 1, making them easier to compare. Specifically, if height data ranges from 160 cm to 180 cm, normalizing converts 160 cm → 0, 170 cm → 0.5, 180 cm → 1, so the minimum becomes 0 and the maximum becomes 1.

(3) Standardization aligns data based on how far each value is from the mean. For example, when test scores from multiple subjects are standardized, even if the mean and spread differ between subjects, the position of each student can be compared.

6. Reshaping Data According to Purpose

(1) Processing that changes the arrangement or grouping of data according to the purpose of the analysis is called **(data reshaping)**. The following three operations are typical.

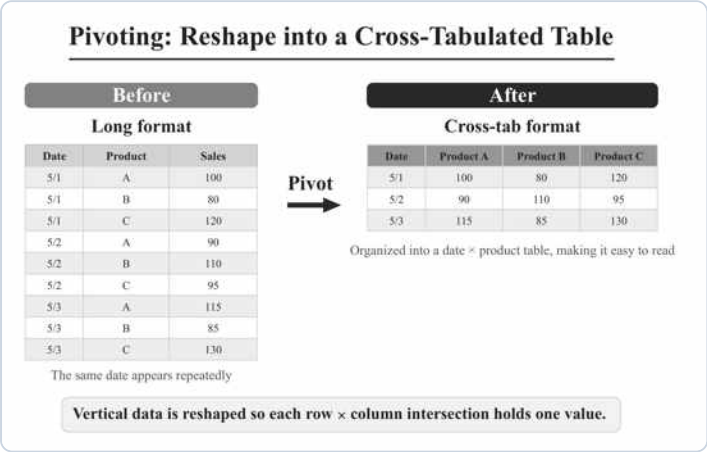
PAUSE & THINK

Height in cm and weight in kg — why might the bigger numbers dominate an analysis?

Normalization rescales variables to a common range.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Operation	Description	Example
(Pivot)	Reshape vertically arranged data into a table of row × column combinations	Reshape data with date, product, and sales arranged vertically into a “date × product” table
(Join)	Connect multiple datasets via a common item	Connect a student roster and a grade sheet by “student ID” into a single table
(Aggregation)	Calculate sums, means, etc. for each group and summarize	Compute monthly totals from daily sales


Pivoting: Reshape into a Cross-Tabulated Table
Before
Long format
Date Product Sales
5/1 A 100
5/1 B 80
5/1 C 120
5/2 A 90
5/2 B 110
5/2 C 95
5/3 A 115
5/3 B 85
5/3 C 130
The same date appears repeatedly
Pivot
After
Cross-tab format
Date Product A Product B Product C
5/1 100 80 120
5/2 90 110 95
5/3 115 85 130
Organized into a date × product table, making it easy to read
Vertical data is reshaped so each row × column intersection holds one value.

Pivoting: reshaping long-format data into a cross-tabulated table

(2) Which operation to perform depends on the purpose of the analysis. For example, if you want to see “monthly sales trends,” you need to aggregate daily sales data by month. It is important to clarify the purpose of the analysis first, and then plan how to reshape the data.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Data cleaning is work performed before data collection.

B The responses to missing values are deletion, imputation, and flagging—three in total.

C All outliers are caused by input or measurement errors, so they must always be deleted.

D Standardization is processing that converts data to a form with mean 0 and standard deviation 1.

(2) Match each description (a–c) with the most appropriate processing from the choices below (A–C).

a Conversion of data to a range from 0 to 1

b Processing that connects different datasets via a common item

c Processing that calculates sums, means, etc. for each group and summarizes

💡 PAUSE & THINK

Is a store's one huge sales day always an error — or could it be real?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

[Choices] **A** Aggregation **B** Join **C** Normalization

✓ Solution

(1)

A Data cleaning is the work of arranging collected data, and is performed after collection. Therefore, ×.

B Correct as the responses to missing values. Therefore, ○.

C Outliers include both abnormal values caused by input errors and values that reflect reality. They are not always deleted. Therefore, ×.

D Correct definition of standardization. Therefore, ○.

(2)

a: Conversion to 0–1 is normalization. **C**

b: Connecting via a common item is join. **B**

c: Summarizing by group is aggregation. **A**

EXAM-STYLE QUESTION

Explain, using "garbage in, garbage out," why data preparation matters as much as the analysis itself. **[6] Refer to:** one cleaning fix and one transformation.

Try

1 Answer the following questions.

1. What is the term for the processing that fills in the missing parts of data with values such as the mean or median?
2. What is the term for the processing that converts data to a range from 0 to 1?
3. What is the term for the processing that converts data to a form with mean 0 and standard deviation 1?
4. What is the term for the operation that connects multiple datasets via a common item?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate response to missing values.

A When there are few missing values and the impact on the whole is small, delete rows containing missing values.

B When missing values are scattered randomly, fill in the missing parts with values such as the mean.

C When the missingness itself has meaning, leave a marker indicating that the value was missing and analyze the data.

D Impute the mean and analyze data on items where most of the data is missing.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) For each of situations (a, b), choose the most appropriate data processing from the choices below (A, B).

a You want to align date data with mixed notations such as “2026/4/1,” “2026-04-01,” and “01 Apr 2026.”

b You want to compare variables with different units and scales together, such as height (cm) and weight (kg).

[Choices] **A** Data type conversion (standardization of date format)
B Normalization

(3) From the following options (A–D), choose the one that most appropriately describes the handling of outliers.

A Outliers are all errors in the data, so delete all of them as soon as they are found.

B For outliers, identify the cause first; correct or delete them if they are due to input errors, and in principle keep them if they reflect reality.

C Outliers do not affect analysis, so no special handling is needed.

D Outliers can always be resolved by replacing them with the mean.

PAUSE & THINK

If most of a column is blank, does filling it with the mean help — or invent data?

Think as an Engineer — Investigate & Decide

Your class collected a survey on daily screen time, but the raw table is messy: some cells are blank, one entry reads “9999” hours, dates are written three different ways, and two rows look identical. You prepare the data for analysis.

☐ **Analyze the raw table as is** ☐ **Clean and transform it first**

1 · Inspect the data. List each problem you can see find; for each, say which category it is (missing value, outlier, duplicate, or format issue) and note in one phrase how it would distort the analysis if left unfixed.

2 · Decide a fix and its risk. For the blank cells, choose deletion, imputation, or flagging and justify it from the proportion of blanks. For “9999,” say how you would check its cause — a typo, a wrong unit (minutes entered as hours), or a code entered for a refused answer — before removing it. Name one respondent whose true answer a careless fix might erase.

3 · Plan the transformation. State the one format conversion you must run so the dates aggregate correctly, and whether you would normalize the screen-time column. **In pairs,** compare cleaning plans and agree on which single fix carries the greatest risk of distorting the results — and why.

Give evidence for your answer.

KEY TERMS

pivot — long rows → row × column table.
join — link datasets by a common item.
aggregation — summarize by group.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? Handle one problem type at a time in order: unify formats first, then check the outlier's cause, then decide the blanks, then remove only true duplicates — and never delete a value before you know why it is there.

🌍 In a New Context

A shop's sales table lists each day's sales for products A, B, and C on separate rows, but the manager wants one row per day with a column per product. Name the reshaping operation needed, and name the operation instead required to get each product's monthly total.

❓ Lesson Question — Answered

The lesson asked why data must be cleaned and transformed before analysis and what choices each fix involves. Collected data is rarely usable as is, and flawed data yields flawed conclusions no matter how careful the later calculation — garbage in, garbage out. Cleaning handles missing values (by deletion, imputation, or flagging, chosen from their proportion and meaning), outliers (judged by cause: correct or delete errors, keep real values), and duplicates (removed only when genuinely duplicated). Transformation makes data comparable and analyzable: data type conversion unifies formats (string to numeric, date formats, notation), normalization rescales to 0–1 and standardization to mean 0 and standard deviation 1 so no variable dominates by size, and reshaping — pivot, join, aggregation — arranges data to fit the question. Each step is a judgment guided by the purpose, not a fixed rule.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

If collected data is used directly in analysis, incorrect conclusions may be drawn. For this reason, (a), which addresses issues, and (b), which arranges the format to fit the purpose of the analysis, are necessary. For missing values, an appropriate response is chosen from deletion, (c), and flagging, depending on the proportion of missing values and the purpose of the analysis. Also, when

⭐ REFLECT & CHALLENGE

Reflect: Why can “filling in” missing data sometimes mislead more than leaving it out? **Challenge:** You have one survey column where 70% of answers are blank. Argue whether to impute, flag, or drop the column, and state the one condition under which your choice would change.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

units or ranges differ greatly between variables, the scale is aligned through (d), which converts data to a range from 0 to 1, or (e), which converts data to a form with mean 0 and standard deviation 1.

1. Fill in the blanks (a)–(e).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that corresponds to data type conversion.

A Filling in missing values with the mean.

B Standardizing notations such as “2026/4/1” and “01 Apr 2026” into “2026-04-01.”

C Aligning the scales of height and weight to between 0 and 1.

D Aggregating daily sales into monthly totals.

(2) From the following options (A–D), choose the one that most appropriately describes normalization and standardization.

A Normalization is processing that converts data to a form with mean 0 and standard deviation 1.

B Standardization is processing that converts data to a range from 0 to 1.

C Normalization is used to align data with different units into a comparable form.

D Standardization is used to enlarge the scale of variables.

(3) For each processing (a–c), indicate which corresponds to pivot, join, or aggregation.

a Connecting a student roster and a grade sheet by “student ID” into a single table.

b Calculating monthly totals from daily sales data.

c Reshaping daily product sales into the form of a “date × product” table.

[Choices] **A** Pivot **B** Join **C** Aggregation

⭐ Key takeaway

Remember:
Clean first (missing values, outliers, duplicates), then transform (convert types, normalize or standardize, reshape by pivot, join, aggregation). Every fix is a judgment set by the purpose — garbage in, garbage out.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Learning Objectives

- Explain what open data is and identify major reliable sources of it.
- Describe how data is obtained through an API, and explain the importance of terms of use and source citation.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Some of the most powerful data in the world is free to use — population figures from the World Bank, global statistics from the UN, competition datasets on Kaggle. But "free" is not the same as "no rules." To rely on open data you need to know where to get it, how to fetch it (a download in a browser, or an API that returns data directly to your program), and how you are allowed to use it. Checking the terms of use and citing your source are what turn borrowed data into trustworthy analysis.

? Guiding question: Where does reliable open data come from, how do programs fetch it, and what rules govern its use?

Point!

1. Using Open Data

(1) To use open data in analysis, it is necessary to understand where to obtain it, how to obtain it, and how to use it. By correctly observing these points, reliable analysis can be performed.

(2) Open data comes from a variety of sources, such as data from international organizations, data from government agencies, and datasets published by researchers and companies. It is important to choose data sources with high reliability that match the purpose of the analysis.

2. Major Open Data Sources

(1) The following are typical, internationally widely used providers of open data.

Provider	Description	Main uses
(World Bank Open Data)	Statistics on economy, population, education, and health for countries around the world	International comparisons and analysis of social/economic trends in each country
(UN Data)	International statistics published by the United Nations and its related agencies	Analysis of global issues such as population, development, and environment
(Kaggle)	A platform that runs data analysis competitions; datasets in various fields are made public	Acquisition of data for machine learning study and research

(2) On these platforms, data can be searched for and downloaded from a browser, and in many cases data can also be obtained through APIs.

3. Obtaining Data via APIs

(1) **(API (Application Programming Interface))**…rules for how pieces of software interact with each other. In web services, APIs are also widely used as a

Key Fact:
Open data comes from official providers (World Bank Open Data, UN Data) and platforms such as Kaggle, whose reliability varies; get it from a browser or through an API (often returning JSON); every dataset has terms of use to follow and a source to cite.

Key Concepts:

- open data
- World Bank Open Data / UN Data / Kaggle
- API / endpoint / JSON
- terms of use
- source citation

KEY TERMS
open data — data made public for anyone to use.
World Bank Open Data — country statistics on economy, population, and more.

Open data is public and free to reuse.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

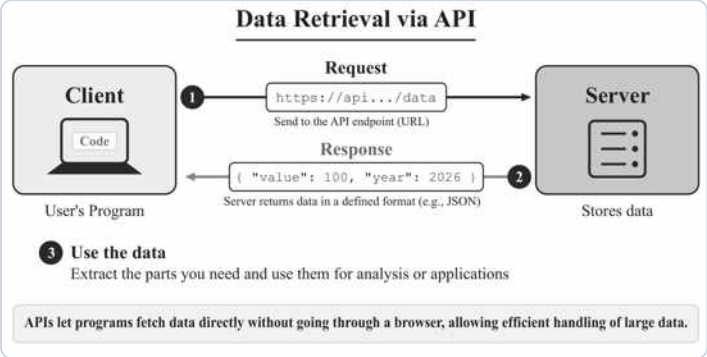
means of obtaining data. By using APIs, necessary data can be obtained directly without going through a browser, allowing large amounts of data to be handled efficiently.

(2) Obtaining data via API proceeds according to the following flow.

① The client (the user's program) sends a request to the API endpoint (the URL that is the destination of the request).

② The server processes the request and returns the requested data in a predetermined format.

③ The client receives the response, extracts the necessary parts, and uses them.


Data Retrieval via API
Client
Code
User's Program
Request
https://api.../data
Send to the API endpoint (URL)
Response
{ "value": 100, "year": 2026 }
Server returns data in a defined format (e.g., JSON)
Server
Stores data
Use the data
Extract the parts you need and use them for analysis or applications
APIs let programs fetch data directly without going through a browser, allowing efficient handling of large data.

Data retrieval via API: request, response, and use

(3) The data format most widely used for data returned via APIs today is **(JSON (JavaScript Object Notation))**. JSON is a notation that represents data as combinations of names and values; it is easy for humans to read and easy to handle in various programming languages.

(4) When using an API, it is important to **(follow the terms of use)**, not send excessive requests in a short time, and confirm the scope of allowed use of the obtained data.

4. Data Use and Source Citation

(1) Even with open data, the scope of free use is set for each dataset. Always confirm the **(terms of use)** of the data before using it.

(2) When using data, it is required to **(cite the source)**. The source generally includes the data provider, the name of the data, the date of acquisition, and the URL.

(3) Citing the source is important both as an indication of the reliability of the data and to allow others to verify the analysis.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A World Bank Open Data publishes statistics on economy, population, education, and so on for countries around the world.

KEY TERMS

API — rules for how software interacts, used to fetch data.
endpoint — the URL a request is sent to.
JSON — a common name/value data format.

An API delivers fresh data straight into analysis.

PAUSE & THINK

"Open" data still has rules — what two things must you check before you use it?

Use data responsibly and credit its source.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

B Using an API, necessary data can be obtained directly without going through a browser.

C As long as data is open data, it can be used freely without checking the terms of use.

D When using data, it is required to cite the provider, the name of the data, the date of acquisition, the URL, and so on as the source.

(2) Match each description (a–c) with the most appropriate open data provider from the choices below (A–C).

a Publishes statistics on economy, population, education, and so on for countries around the world.

b Provides international statistics published by the United Nations and its related agencies.

c A platform that runs data analysis competitions.

[Choices] **A** UN Data **B** World Bank Open Data **C** Kaggle

PAUSE & THINK

In an API call, who sends the request and who returns the data?

✓ Solution

(1)

A Correct description of World Bank Open Data. Therefore, ○.

B Correct description of an API. Therefore, ○.

C Even for open data, the terms of use differ for each dataset, so they must always be checked before use. Therefore, ×.

D Correct content of source citation. Therefore, ○.

(2)

a: Statistics on countries around the world is World Bank Open Data. **B**

b: Data from the United Nations and its related agencies is UN Data. **A**

c: The platform that runs data analysis competitions is Kaggle. **C**

EXAM-STYLE QUESTION

Explain why an analyst must check the terms of use and cite the source even when the data is free and open. **[6] Refer to:** your right to use the data, reliability, and letting others verify.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the name of the World Bank's open data portal that publishes statistics on economy, population, education, and so on for countries around the world?
2. What is the term for the rules for how pieces of software interact with each other, which is also widely used in web services for obtaining data?
3. What is the name of the data format widely used when data is returned from an API?
4. What is the term for what should be cited when using data, both to indicate reliability and to allow verification by others?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the flow of obtaining data via an API.

A The client sends a request to the server, and the server returns the data in a format such as JSON.

B The server one-sidedly continues to send data to the client.

C Clients exchange data directly with each other.

D The server only stores data; clients cannot retrieve it.

(2) For each of purposes (a, b), choose the most appropriate open data provider from the choices below (A–C).

a You want to compare GDP across countries.

b You want to find datasets from various fields for machine learning study.

[Choices] **A** UN Data **B** World Bank Open Data **C** Kaggle

(3) From the following options (A–D), choose the one that most appropriately describes the use of open data.

A Open data can be used freely, so the terms of use do not need to be checked.

B Source citation is required only for data published for a fee.

C Even when using open data, it is required to cite the provider, the name of the data, the date of acquisition, the URL, and so on as the source.

D For data obtained via an API, there is no need to follow the terms of use.

PAUSE & THINK

Which provider fits which purpose — and would you use a browser or an API?

Think as an Engineer — Investigate & Decide

For a school project you want to show how Internet access in Egypt has changed over the past ten years compared with two neighboring countries.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

You must choose an open data source and a way to obtain it.

☐ **Take the first dataset you find** ☐ **Choose a reliable source and follow its rules**

1 · Inspect two sources. Look up two candidate providers for this exact question (for example World Bank Open Data and UN Data). For each, note whether it actually holds Internet-access figures by country and year, and how current the latest year is.

2 · Analyze fit and rules. Compare the two on a point the lesson table does not list — such as how far back the data goes, or whether its terms of use allow you to republish a chart in a public report. Name one risk of picking the wrong source here.

3 · Decide. Choose one provider and one fetch method (browser download or API) and justify both from what you found. Write the source citation you would include (provider, data name, date obtained, URL). **In pairs,** compare choices and agree on the more defensible source.

Give evidence for your answer.

Stuck? Start from the question, not the source: list exactly what fields you need (country, year, Internet-access rate), then keep only the providers that actually carry all three — and check each one's terms before you commit.

KEY TERMS

terms of use — the rules set for each dataset.

source citation — provider, name, date, URL.

🌍 In a New Context

A weather app on a phone shows today's forecast by fetching it from a public weather service. Describe the request-and-response steps between the app and the server, and name the data format the server most likely returns.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

❓ Lesson Question — Answered

The lesson asked where reliable open data comes from, how programs fetch it, and what rules govern its use. Reliable open data comes from trusted providers chosen to fit the purpose — World Bank Open Data for country economic and social statistics, UN Data for United Nations international statistics, and Kaggle for datasets used in machine learning study. It can be downloaded through a browser or fetched by a program through an API: the client sends a request to an endpoint, and the server returns the data in a set format, most often JSON, which the client then uses. Because open data still comes with rules, every dataset carries terms of use that must be followed — including not overloading an API with requests — and every use requires citing the source (provider, data name, date of acquisition, URL), which signals reliability and lets others verify the analysis.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

To use open data in analysis, it is necessary to understand where to obtain it, how to obtain it, and how to use it. Statistics on economy, population, education, and so on for countries around the world can be obtained from sources such as (a), operated by the World Bank. Also, in addition to downloading individually from a browser, by using (b), necessary data can be obtained directly in formats such as (c) in response to a request. When using open data, the (d) determined for each dataset must always be checked, and when data is used, the provider, the name of the data, the date of acquisition, the URL, and so on must be cited as the (e).

1. Fill in the blanks (a)–(e).

⭐ REFLECT & CHALLENGE

Reflect: Think of a chart you saw online with no source. Did you trust it — and what one line would have changed that?
Challenge: You may cite only one source line for a chart built from a World Bank dataset you downloaded today. Write that citation line, and say which single element you would never drop and why.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes Kaggle.

A A portal operated by the World Bank that publishes statistics for countries around the world.

B A site for international statistics operated by the United Nations.

C A platform that runs data analysis competitions, on which datasets in various fields are made public.

D A system that automatically records browsing histories of websites.

(2) From the following options (A–D), choose the one that is NOT an appropriate description of source citation.

A Source citation is important as an indication of the reliability of the data.

B Source citation is important to allow others to verify the analysis.

C Sources include the data provider, the name of the data, the date of acquisition, the URL, and so on.

D When using open data, it is not necessary to cite the source.

(3) From the following options (A–D), choose the one that is NOT an appropriate point of caution when using an API.

A Use the API in accordance with the terms of use determined by the provider.

B Avoid sending excessive requests in a short time.

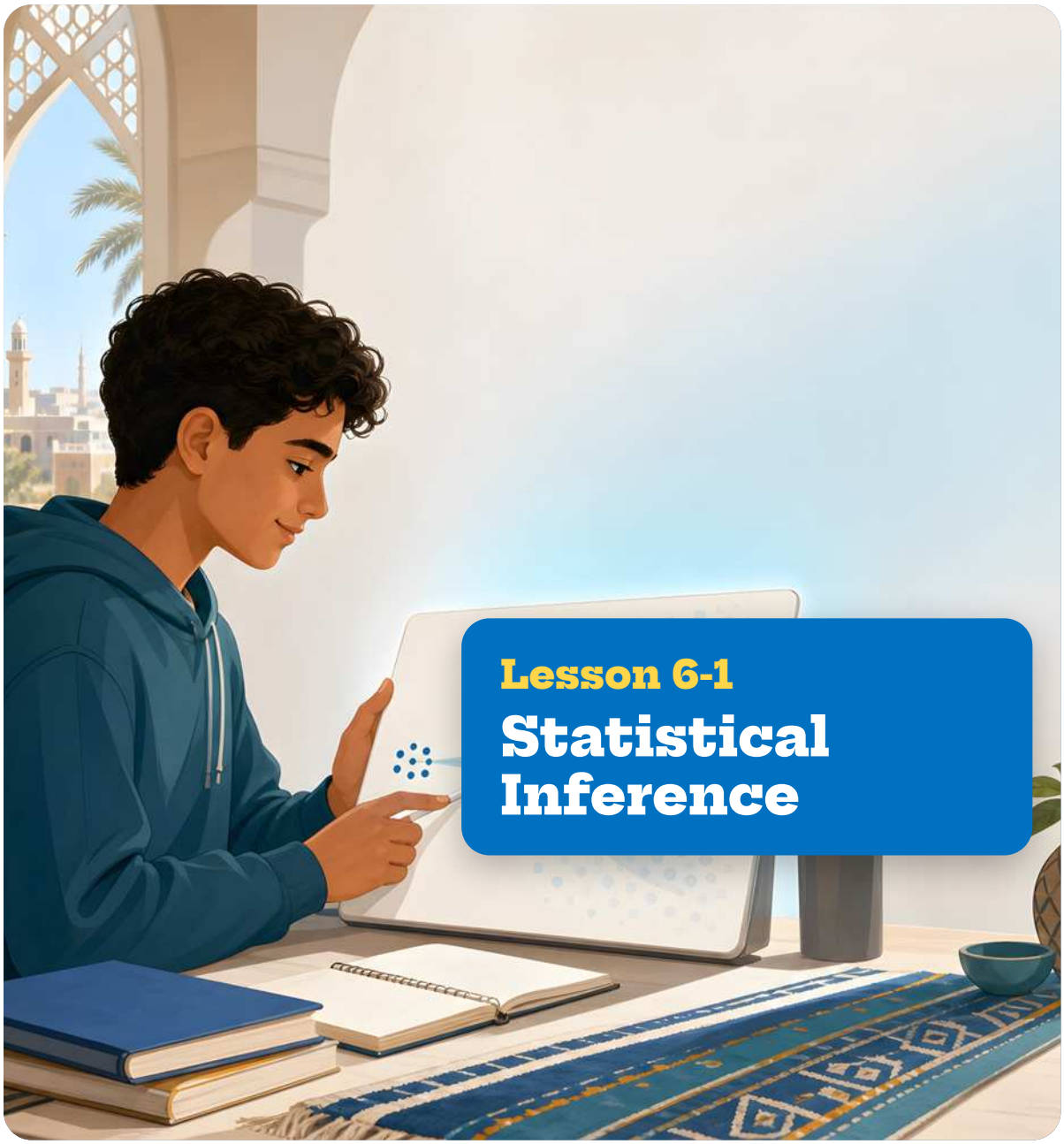
C Confirm the scope of allowed use of the obtained data in advance.

D Because of how APIs work, they may be used without checking the terms of use.

⭐ Key takeaway

Remember:
Pick a reliable open data source for your purpose (World Bank Open Data, UN Data, Kaggle), fetch by browser or API (client requests an endpoint; server returns JSON), and always follow the terms of use and cite the source.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Learning Objectives

- Explain the difference between point estimation and interval estimation, including the meaning of a confidence level.
- Describe the procedure of hypothesis testing, and explain the cautions needed when interpreting its results.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

You can never survey everyone, so you study a sample and reason back to the whole population — that is statistical inference. It answers two kinds of question: "about how big?" (estimation) and "is this real, or just chance?" (hypothesis testing). But because a sample is only part of the whole, every answer carries uncertainty. The skill is not to hide that uncertainty but to measure it — with confidence intervals, p-values, and significance levels — and to remember that "statistically significant" is not the same as "important."

? Guiding question: How can we reason from a sample to a whole population, and how do we measure the uncertainty in that reasoning?

📌 Point!

1. The Statistical Inference Stage

(1) A sample is drawn from the entire group that is the target of analysis (the population) and surveyed, and the properties of the population are then inferred from the results of the sample. Inferring the properties of the population from the sample in this way is called **(statistical inference)**.

(2) Statistical inference can be broadly divided into the following two approaches.

Approach	Description	Example
(Estimation)	Infer the properties of the population in numerical terms	Infer the population mean height from the sample mean height
(Hypothesis testing)	Determine whether a hypothesis about the population is correct	Determine whether a new learning method is more effective than the conventional one

(3) In either case, since the sample is only part of the population, inference always involves uncertainty. In statistical inference, judgment is made while considering the degree of that uncertainty.

2. Point Estimation and Interval Estimation

(1) **(Point estimation)**…estimating the properties of the population as a single value calculated from the sample. *e.g.,* If the sample mean height is 168 cm, then the population mean height is also estimated to be 168 cm.

(2) **(Interval estimation)**…estimating that the properties of the population fall within a certain range (interval). *e.g.,* The population mean height is estimated to be in the range of 166 cm to 170 cm.

(3) The interval used in interval estimation is called the **(confidence interval)**, and how often such intervals contain the population value over repeated surveys is called the **(confidence level)**. Levels of 95% or 99% are common.

Key Fact:
Statistical inference reasons from a sample to a population. Estimation gives a value (point) or a range (interval, with a confidence level); hypothesis testing weighs a null against an alternative using a p-value and a significance level. All of it carries uncertainty — and significance is not the same as practical importance.

Key Concepts:

- statistical inference
- point / interval estimation
- confidence interval & level
- null / alternative hypothesis
- p-value & significance level
- Type I / II errors

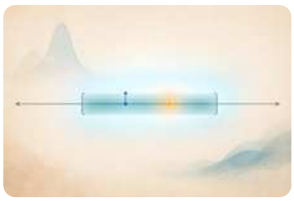
KEY TERMS
statistical inference — reasoning from a sample to the population.
estimation — inferring the population in numbers.
hypothesis testing — judging a claim about the population.

💡 PAUSE & THINK
A single value or a range — which estimate shows you how uncertain it is?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

e.g., Suppose the population mean height is estimated by sample survey, and a confidence interval of "166 cm to 170 cm" is obtained at a 95% confidence level. The "95%" means that if the same survey method is repeated 100 times, about 95 of the resulting confidence intervals will contain the true population mean height.

(4) Point estimation is easy to understand because it is shown as a single value, but the uncertainty of the estimate is invisible. Interval estimation has width, so care is needed in interpreting the result, but it has the advantage of also showing the degree of uncertainty.



A point estimate is one value; an interval is a range.

3. The Idea of Hypothesis Testing

(1) Hypothesis testing…a procedure for determining whether a hypothesis about the population is acceptable based on the data at hand.

(2) In hypothesis testing, the following two hypotheses are set up.

Hypothesis	Description
(Null hypothesis) (H0)	A hypothesis serving as the starting point of the test, such as "no difference" or "no relationship"
(Alternative hypothesis) (H1)	The hypothesis adopted when the null hypothesis is rejected, such as "there is a difference" or "there is a relationship"

(3) In hypothesis testing, the probability that data collection could result in the data at hand is evaluated under the assumption that "there is no difference or relationship." The reason for first assuming the hypothesis one wants to deny (the null hypothesis) is that, rather than directly showing "there is a difference or relationship," it is logically easier to judge this by showing that, under the assumption "there is no difference or relationship," the data at hand would be unlikely to be obtained. *e.g.,* It is difficult to directly show "the new learning method is effective." Instead, one assumes "there is no effect" and then examines how low the probability is of obtaining a difference like the experimental result at hand, and if that probability is sufficiently small, one judges that "no effect" is unlikely.

(4) Based on the evaluation, the null hypothesis is judged to be either **(rejected (denied))** or **(not rejected)**.

Testing weighs evidence against the null hypothesis.

KEY TERMS
null hypothesis (H0) — the "no difference" starting point.
alternative hypothesis (H1) — adopted if H0 is rejected.
p-value — chance of data at least this extreme if H0 holds.

4. The Procedure of Hypothesis Testing

(1) Hypothesis testing follows this flow.

① Set the null hypothesis H0 and the alternative hypothesis H1

② Decide the **(significance level)** (5% = 0.05 and 1% = 0.01 are commonly used)

③ Compute the **(p-value)** from the data at hand

④ Compare the p-value with the significance level and make a judgment

(2) p-value…assuming the null hypothesis is correct, the probability of obtaining a result as extreme as, or more extreme than, the data at hand. The

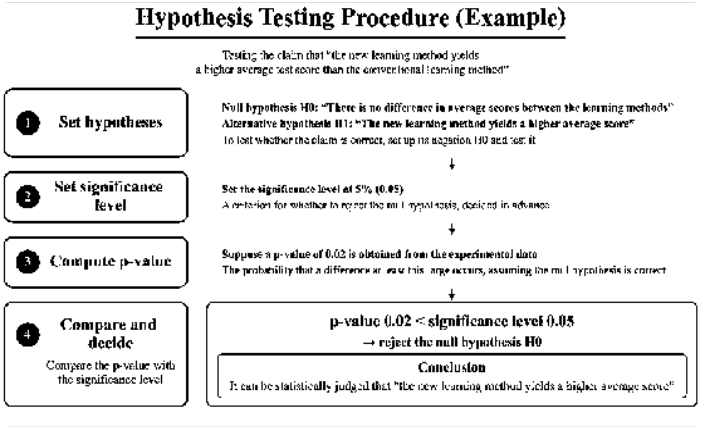
D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

smaller the p-value, the less likely it is that the data would be obtained under the null hypothesis.

(3) Significance level…the threshold value used to decide whether to reject the null hypothesis.

(4) Criteria for judgment

Condition	Judgment
p-value < significance level	Reject the null hypothesis and adopt the alternative hypothesis (statistically significant)
p-value ≥ significance level	Do not reject the null hypothesis (cannot be said to be statistically significant)


Hypothesis Testing Procedure (Example)
Testing the claim that "the new learning method yields a higher average test score than the conventional learning method"
1 Set hypotheses
Null hypothesis H0: "There is no difference in average scores between the learning methods"
Alternative hypothesis H1: "The new learning method yields a higher average score"
To test whether the claim is correct, set up the negation H0 and test it
2 Set significance level
Set the significance level at 5% (0.05)
A criterion for whether to reject the null hypothesis, decided in advance
3 Compute p-value
Suppose a p-value of 0.02 is obtained from the experimental data
The probability that a difference at least this large occurs, assuming the null hypothesis is correct
4 Compare and decide
Compare the p-value with the significance level
p-value 0.02 < significance level 0.05
→ reject the null hypothesis H0
Conclusion
It can be statistically judged that "the new learning method yields a higher average score"

The hypothesis testing procedure worked through an example

5. Major Hypothesis Testing Methods

(1) The test method used differs depending on the purpose of the analysis and the type of data. Major examples include the following.

Test	Main use	Example
(t-test)	Examine whether there is a difference in the means of two groups	Examine whether students' average test scores differ between a certain learning method and the conventional method
(Chi-squared test)	Examine whether there is a relationship between two categorical variables (data expressed by classification, such as gender, occupation, or preference)	Examine, from survey results, whether there is a relationship between gender and subject preference

6. Cautions Regarding Inference

(1) In statistical inference, there is always a possibility of judgment errors. There are two types of errors.

Type of error	Description
(Type I error)	Rejecting the null hypothesis when it is actually correct
(Type II error)	Not rejecting the null hypothesis when it is actually incorrect

PAUSE & THINK

If the p-value is 0.02 and the significance level is 0.05, do you reject the null or not?

PAUSE & THINK

Which test compares two group means, and which checks a relationship between categories?

KEY TERMS

Type I error — rejecting a true null hypothesis.
Type II error — keeping a false null hypothesis.
trade-off — lowering one error raises the other.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(2) Lowering the significance level reduces the Type I error, but tends to increase the Type II error. The two have a **(trade-off relationship)**.

(3) Care is needed in interpreting the p-value. The following misunderstandings must be avoided.

① The p-value is not “the probability that the null hypothesis is correct.”

② A small p-value does not mean “the effect is large.”

③ The comparison of the p-value with the significance level is not an absolute correctness criterion but a guide for judgment. The significance level (e.g., 0.05 or 0.01) is a value chosen according to the purpose of the analysis or the conventions of the field; falling below this value does not mean the result is certainly correct.

(4) Being statistically significant and being practically important are different. *e.g.,* Suppose an experiment compares two designs of a website, with click-rate data from 100,000 people split evenly. The old design is clicked by 0.4% of visitors, the new one by 0.5% — only 0.1 percentage points higher. Because the data set is large, this difference is likely to be statistically significant. However, whether that 0.1-point difference leads to actual profit or better user experience needs to be judged from a separate perspective.

(5) Statistical inference is a powerful means of drawing conclusions from data, but it is important to understand that it is a tool for judgment that includes uncertainty, and the results should be interpreted appropriately.

Type I is a false alarm; Type II is a miss.

PAUSE & THINK

With 100,000 people, a tiny difference can be “significant.” Does that make it important?

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Inferring the properties of the population from the sample is called statistical inference.

B Point estimation is a method that estimates the properties of the population within a certain range.

C In hypothesis testing, the probability that data collection could result in the data at hand is evaluated under the assumption that the null hypothesis is correct.

D If a result is statistically significant, the difference is also necessarily important in practice.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A hypothesis serving as the starting point of the test, such as “no difference” or “no relationship”

b The probability of obtaining a result as extreme as, or more extreme than, the data at hand, assuming the null hypothesis is correct

c The threshold value for deciding whether to reject the null hypothesis

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

[Choices] **A** p-value **B** Significance level **C** Null hypothesis

✓ Solution

(1)

A Correct definition of statistical inference. Therefore, ○.

B Estimating within a range is interval estimation. Point estimation estimates a single value. Therefore, ×.

C Correct description of the idea of hypothesis testing. Therefore, ○.

D Statistical significance and practical importance are different; with a large amount of data, even a tiny difference can become significant. Therefore, ×.

(2)

a: A hypothesis serving as the starting point of the test is the null hypothesis. **C**

b: The probability of obtaining the data under the null hypothesis is the p-value. **A**

c: The threshold value for rejection is the significance level. **B**

Try

1 Answer the following questions.

1. What is the term for inferring the properties of the population from the sample?
2. What is the term for the method of estimating that the properties of the population fall within a certain range?
3. What is the term for the interval used in interval estimation?
4. What is the term for a typical hypothesis testing method that examines whether there is a relationship between two categorical variables?

EXAM-STYLE QUESTION

Explain why researchers test a null hypothesis they hope to disprove, rather than proving the effect directly. **[6] Refer to:** what a small p-value shows.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

2 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the procedure of hypothesis testing.

A Set the null and alternative hypotheses → decide the significance level → compute the p-value → make a judgment

B Compute the p-value → set the null hypothesis → decide the significance level → make a judgment

C Decide the significance level → adopt the alternative hypothesis → compute the p-value → set the null hypothesis

D Make a judgment → set the null hypothesis → decide the significance level → compute the p-value

(2) Suppose that, in a hypothesis test with the significance level set at 0.05 (5%), the p-value is 0.03. From the following options (A–D), choose the one that most appropriately describes the judgment in this case.

A Reject the null hypothesis and adopt the alternative hypothesis.

B Do not reject the null hypothesis.

C Reject the alternative hypothesis.

D Change the significance level to 0.01 and recalculate.

(3) For each of analysis purposes (a, b), choose the most appropriate test method from the choices below (A, B).

a You want to examine whether students' average test scores differ between two teaching methods.

b You want to examine whether there is a relationship between gender and the category of products purchased.

[Choices] **A** Chi-squared test **B** t-test

(4) From the following options (A–D), choose the one that most appropriately describes the p-value.

A The p-value is the probability that the null hypothesis is correct.

B The smaller the p-value, the larger the effect is.

C The p-value is the probability of obtaining a result as extreme as, or more extreme than, the data at hand, assuming the null hypothesis is correct.

D p-value < 0.05 is an absolute criterion of correctness.

Think as an Engineer — Investigate & Decide

A school club claims its new online quiz tool raises students' average score on weekly tests. You are asked to judge the claim fairly using a hypothesis test.

☐ **Trust the club's own summary** ☐ **Frame and test the claim properly**

PAUSE & THINK

p-value 0.03 against a 0.05 threshold — below or above? So what do you do?

KEY TERMS

significance level — the threshold for rejecting H0.
statistically significant — p below the threshold.
practically important — big enough to matter in real life.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1 · Frame the hypotheses. The club reports its tool group averaged 74 on the last quiz versus 70 for a control group of the same size. Write the null and alternative hypotheses for this exact claim in your own words, state what data you would still need to actually run the test (beyond these two means), and say which significance level you would set (0.05 or 0.01) and why.

2 · Analyze the risks. Describe, in this context, what a Type I error and a Type II error would each mean for the club, and say which one you would most want to avoid here — a judgment the lesson's tables do not make for you.

3 · Decide and interpret. Suppose the test returns a p-value of 0.04. State your decision against the significance level you chose in step 1 (note it would differ at 0.05 versus 0.01). Then, if the result is significant, explain why even a significant result might still not justify switching every class to the tool. **In pairs,** compare interpretations and agree on what evidence beyond the p-value you would still want.

Give evidence for your answer.

Stuck? Start from what you are trying to disprove: the null is always “no difference.” Then ask what it would cost to reject it wrongly (Type I) versus to miss a real effect (Type II) — that cost is what should guide your significance level.

🌍 In a New Context

A shop tests two checkout-button colors with data from 200,000 visitors and finds the new color's click rate is 0.05 percentage points higher, with a p-value of 0.01 at the 5% significance level. State whether the result is statistically significant, and explain why the shop might still decide not to change the button.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

❓ Lesson Question — Answered

The lesson asked how we reason from a sample to a whole population and how we measure the uncertainty in that reasoning. Because a sample is only part of the population, we use statistical inference — either estimation, which reports the population as a single value (point estimation) or as a range (interval estimation) whose confidence level says how often such intervals would contain the true value, or hypothesis testing, which sets a null hypothesis ("no difference") against an alternative and compares a p-value with a chosen significance level to decide whether to reject the null. The uncertainty is never removed: it is measured and reported. Judgment can fail as a Type I error (rejecting a true null) or a Type II error (keeping a false null), and lowering one raises the other. And a result that is statistically significant — especially with very large samples — is not automatically important in practice, so results must be interpreted, not just computed.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

In statistical inference, the properties of the population are inferred from the sample. Estimating the properties of the population as a single value is called (a), and estimating within a certain range is called (b). Also, the procedure of judging a hypothesis about the population is called (c). In hypothesis testing, the (d)—such as "no difference" or "no relationship"—and the opposite (e) are set up, and a judgment is made by comparing the (f) computed from the data at hand with the predetermined (g).

1. Fill in the blanks (a)–(g).

⭐ REFLECT & CHALLENGE

Reflect: Why is a 95% confidence level a statement about the method, not about one particular interval?

Challenge: A study of 500,000 users reports a "highly significant" effect ($p < 0.001$) whose size is 0.02 percentage points. Argue whether you would act on it, and name the one extra piece of information that would change your mind.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes confidence intervals and confidence levels.

A A confidence interval with a 95% confidence level means that the true population value is contained in that interval with a probability of 95%.

B The higher the confidence level, the narrower the confidence interval.

C A confidence interval with a 95% confidence level means that, if the same survey method is repeated, about 95% of the resulting confidence intervals will contain the true population value.

D The confidence level is a value used as the criterion for judgment in hypothesis testing.

(2) From the following options (A–D), choose the one that is NOT an appropriate description of errors in hypothesis testing.

A A Type I error is rejecting the null hypothesis when it is actually correct.

B A Type II error is not rejecting the null hypothesis when it is actually incorrect.

C Lowering the significance level can simultaneously reduce both Type I and Type II errors.

D Type I and Type II errors have a trade-off relationship.

(3) From the following options (A–D), choose the one that most appropriately describes the interpretation of statistical inference.

A A statistically significant difference is necessarily an important difference in practice.

B With a large amount of data, even a tiny difference can become statistically significant.

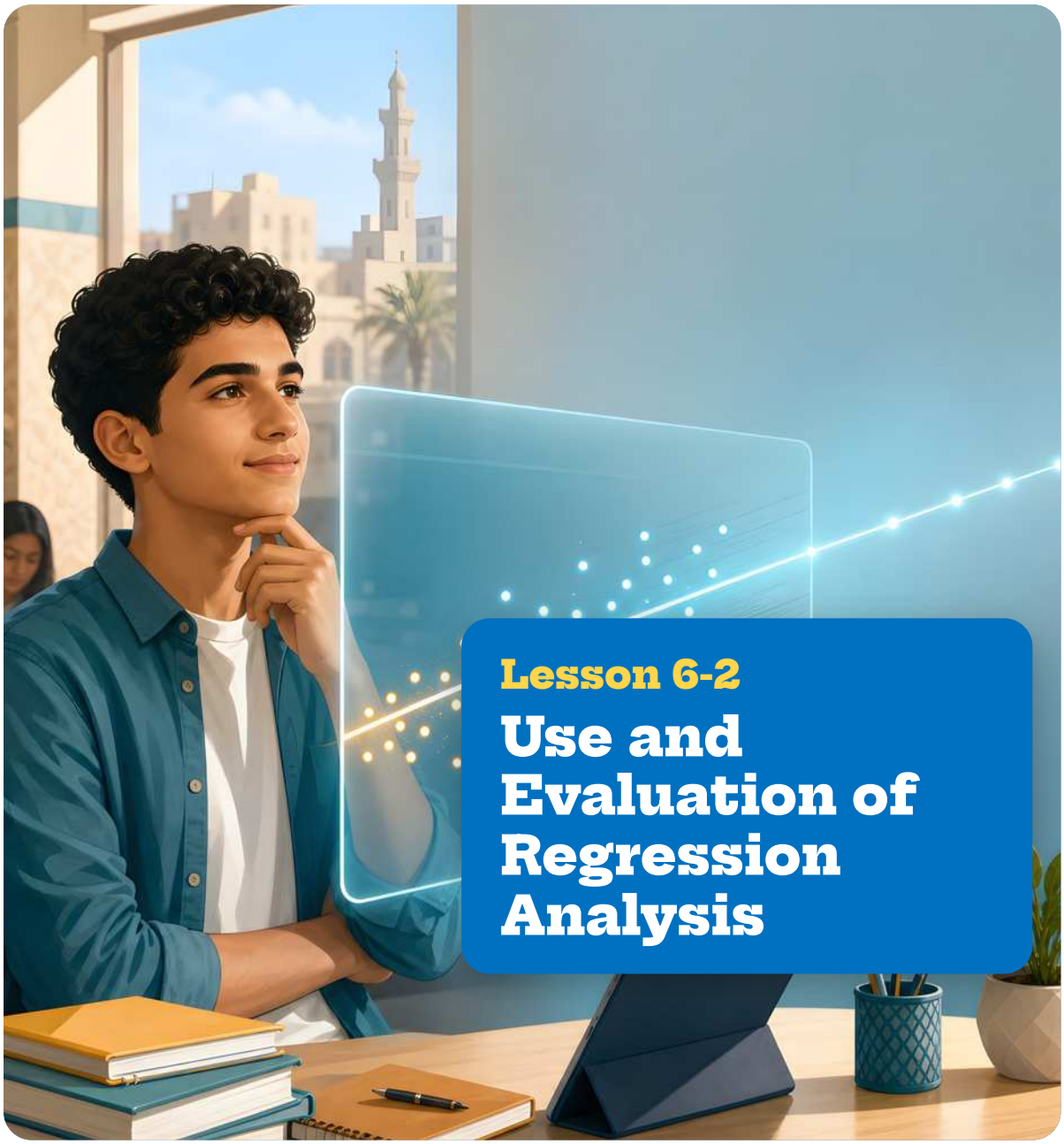
C If the p-value falls below 0.05, the conclusion is certainly correct.

D Statistical inference can completely remove uncertainty.

⭐ Key takeaway

Remember:
Estimate with a point or an interval (confidence level = how often the method captures the truth); test by comparing a p-value with a significance level to reject or keep the null. Watch Type I / II errors, and never confuse statistical significance with practical importance.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 6-2

Use and Evaluation of Regression Analysis

Learning Objectives

- Explain the difference between simple and multiple regression, and what a regression model represents.
- Evaluate a regression model using R^2, RMSE, and MAE, and explain why a good fit does not prove causation.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Data often hides a relationship you can turn into a prediction: the warmer the day, the more ice cream sells. Regression analysis writes that relationship as an equation, so you can put in tomorrow's forecast temperature and predict tomorrow's sales. Yet a prediction equation is only a model — a simplified picture — and a careful analyst asks two more questions: how accurate is it (using R^2, RMSE, and MAE), and does "fits well" really mean "causes"? A line that fits the data is not proof of cause.

? Guiding question: How does regression turn data into a prediction, and how do we judge whether that prediction can be trusted?

Point!

1. The Regression Analysis Stage

(1) **(Regression analysis)**…a method that expresses the relationship between variables in the data as an equation, and uses it for prediction and simulation. For example, from past data on temperature and sales, one can express how sales change when temperature rises as an equation, and put tomorrow's forecast temperature into that equation to predict sales.

(2) In regression analysis, the variable that is the result is called the **(response variable) (dependent variable)**, and the variable used to explain or predict it is called the **(explanatory variable)**. In the temperature-and-sales example, sales are the response variable and temperature is the explanatory variable.

(3) The equation obtained from data in regression analysis is called the **(model)**. A model is a simplified representation of a real relationship—for example, putting the trend "the higher the temperature, the higher the sales" into a format that can be calculated.

(4) Both creating a model and evaluating whether the model can actually be used are important. A model that is only created is not useful; only by confirming its accuracy and limitations does it lead to practical use.

2. Simple and Multiple Regression

(1) **(Simple regression analysis)**…a method of predicting the response variable using one explanatory variable.

e.g., Predict ice cream sales (y) from temperature (x) alone. The prediction equation is expressed as a straight line in the format **$y = ax + b$**, and this line is called the **(regression line)**. a and b are determined by the **(least-squares method)** (a method that determines the line so as to minimize the sum of the squares of the discrepancies between the actual values and the regression line, each called a **(residual)**).

(2) **(Multiple regression analysis)**…a method of predicting the response variable using multiple explanatory variables. In real data, results are rarely determined by a single cause, and multiple regression analysis, which can consider several factors simultaneously, is suited to more realistic prediction.

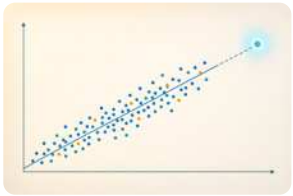
Key Fact:
Regression analysis expresses the relationship between an explanatory variable (the input used to make the prediction) and a response variable (the result) as a model — a line $y = ax + b$ for simple regression, or a sum of weighted variables for multiple regression. Judge the model with R^2, RMSE, and MAE, and never mistake a good fit for proof of causation.

Key Concepts:

- regression analysis
- response / explanatory variable
- simple / multiple regression
- regression line, least-squares
- R^2, RMSE, MAE
- correlation versus causation

KEY TERMS

regression analysis — expressing a relationship as an equation for prediction.
response variable — the result (dependent).
explanatory variable — the input used to make the prediction.



Fit a line to data, then use it to predict within that range.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

e.g., Predict housing prices from a house's size, age, and distance from a transit station (multiple factors). The prediction equation is expressed as the sum of multiple explanatory variables, each multiplied by a coefficient (a numerical value representing the strength of its influence), plus an intercept. For example, when there are two explanatory variables, the format becomes $\boldsymbol{y = a_1x_1 + a_2x_2 + b}$ (x_1 and x_2 are the explanatory variables, a_1 and a_2 are their coefficients, and b is the intercept). When there are three or more explanatory variables, terms are added in the same way.

(3) The following differences exist between simple and multiple regression.

Aspect	Simple regression	Multiple regression
Number of explanatory variables	One	Two or more
Characteristics of the model	Simple and easy to interpret	Can express complex relationships
Suitable situations	When you want to capture the relationship between two variables concisely	When the response variable is influenced by multiple factors

3. Evaluation Indicators for Models

(1) To judge how useful a model created by regression analysis is, **(evaluation indicators)** are used. Three typical indicators are as follows.

Indicator	Description (concept of calculation)	Interpretation
(R^2 (coefficient of determination))	Expresses, on a 0–1 scale, how well the model explains the variability in the data	The closer to 1, the better the variation in the data is explained
(RMSE (root mean square error))	The square root of the mean of the squared discrepancies between predicted and actual values	The smaller the value, the higher the prediction accuracy; expressed in the same units as the response variable
(MAE (mean absolute error))	The mean of the absolute values of the discrepancies between predicted and actual values	The smaller the value, the higher the prediction accuracy; less affected by extremely large discrepancies

(2) Both RMSE and MAE indicate the size of prediction errors, but RMSE tends to weight large discrepancies more heavily. On the other hand, MAE evaluates discrepancies on average, so it is less affected by outliers. They are used differently according to the purpose of the analysis.

4. Cautions When Using Regression Models

(1) Regression analysis is a widely applicable method, but misuse can lead to incorrect conclusions. Pay particular attention to the following points.

PAUSE & THINK

One cause or several? Which kind of regression would house prices need?

Multiple regression combines several inputs.

KEY TERMS

R^2 — how well the model explains the variability in the data (0–1).
RMSE — error size, weights big misses more.
MAE — average error, less swayed by outliers.

PAUSE & THINK

A prediction line fits the data well — does that prove one variable causes the other?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

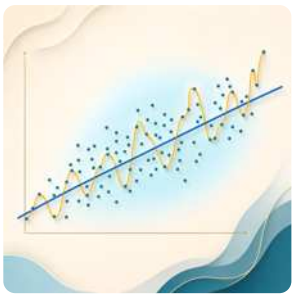
① Do not confuse correlation with causation: a prediction equation that fits does not establish it. Causation needs temporal precedence, exclusion of other factors' influence, and theoretical support.

② A model is not finished once created: Confirm accuracy with evaluation indicators, verify validity in light of the purpose, and do not apply it outside the range of the data it was fitted to.

③ Do not judge a model by R^2 alone: Make judgments together with other indicators.

④ Adding more explanatory variables does not necessarily improve a model: Appropriate variables must be chosen according to the purpose.

(2) The results of regression analysis should not be treated as absolute truth as is, but as one perspective that the data offers. Interpreting results appropriately and using them in real-world judgment is important.



An overfit model chases noise, not the trend.

PAUSE & THINK

Which error measure would you trust more when a few huge outliers exist?

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Multiple regression analysis is a method of predicting the response variable using one explanatory variable.

B R^2 (coefficient of determination) is an indicator showing how well the model explains the variability in the data.

C The equation obtained from data in regression analysis is called the model, and is a simplified representation of a real relationship.

D If the prediction equation fits, it can be said that there is a causal relationship between the explanatory variable and the response variable.

(2) Match each description (a–c) with the most appropriate indicator from the choices below (A–C).

a The mean of the absolute values of the discrepancies between predicted and actual values

b The square root of the mean of the squared discrepancies between predicted and actual values

c How well the model explains the variability in the data

[Choices] **A** RMSE **B** R^2 **C** MAE

✓ Solution

(1)

A Predicting with one explanatory variable is simple regression analysis. Multiple regression analysis uses multiple explanatory variables. Therefore, ×.

B Correct description of R^2. Therefore, ○.

C Correct definition of a model. Therefore, ○.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

D Even if the prediction equation fits, that alone does not establish causation. Therefore, ×.

(2)

a: The mean of absolute discrepancies is MAE. **C**

b: The square root of the mean of squared discrepancies is RMSE. **A**

c: An indicator that explains the variability in the data is R^2. **B**

Try

1 Answer the following questions.

1. What is the term for the regression analysis that predicts the response variable using multiple explanatory variables?
2. What is the term for the indicator showing how well the model explains the variability in the data?
3. In regression analysis, what are the variables called that are the result and the input used to predict it, respectively?
4. What is the term for “the discrepancy between actual values and the regression line” used in simple regression analysis?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the difference between simple and multiple regression.

A Simple regression can express complex relationships, while multiple regression is simple.

B Simple regression uses multiple explanatory variables, and multiple regression uses one explanatory variable.

C Simple regression uses one explanatory variable, and multiple regression uses multiple explanatory variables.

D There is no difference between simple and multiple regression in the number of explanatory variables; only the calculation method differs.

(2) For each of analysis purposes (a, b), choose the most appropriate regression analysis method from the choices below (A, B).

a You want to predict ice cream sales from temperature (one variable).

b You want to predict housing prices from a house's size, age, and distance from a transit station (multiple variables).

[Choices] **A** Simple regression analysis **B** Multiple regression analysis

EXAM-STYLE QUESTION

Explain why a regression model with a high R^2 can still be a poor basis for claiming that one variable causes another. **[6] Refer to:** correlation versus causation.

PAUSE & THINK

One variable in, or several? That is the whole difference.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that most appropriately describes the interpretation of regression analysis results.

A If a prediction equation is obtained from regression analysis, the equation necessarily indicates a causal relationship.

B The results of regression analysis should be treated as one perspective that the data offers, and whether causation has been demonstrated must be carefully evaluated.

C If R^2 is high, the model is necessarily a good model.

D If the prediction equation fits, one may conclude that the explanatory variable is the cause of the response variable.

(4) From the following options (A–D), choose the one that is NOT an appropriate description of evaluation indicators for regression models.

A R^2 (coefficient of determination) takes values between 0 and 1; the closer to 1, the better the variation in the data is explained.

B RMSE is the square root of the mean of squared discrepancies between predicted and actual values, expressed in the same units as the response variable.

C MAE is the mean of the absolute values of the discrepancies between predicted and actual values, and is less affected by extremely large discrepancies.

D If the value of R^2 is close to 1, the model is necessarily a good model without needing to check other indicators.

PAUSE & THINK

A high R^2 alone — enough to trust a model, or do you need RMSE and MAE too?

Think as an Engineer — Investigate & Decide

Your school canteen wants to predict its daily cold-drink sales so it can order the right amount. You plan a regression model.

☐ **Trust one obvious cause** ☐ **Model it and test the model**

1 · Set up the model. Name the response variable, then list the explanatory variables you would try (for example the day's temperature, the number of students present, and whether it is an exam day). Decide whether this is a simple or a multiple regression, and say why.

2 · Judge and question it. Choose which evaluation indicator you would report (R^2, RMSE, or MAE) and justify it for this purpose. Then name one confounder (a confounding variable) — a hidden third factor that could make two variables move together without one causing the other — that could mislead you here.

3 · Decide and interpret. Suppose your model gives a high R^2. State whether that alone would let you claim "hot days cause higher sales," and what else you would need. **In pairs,** compare models and agree on the single most useful explanatory variable to keep.

Give evidence for your answer.

KEY TERMS

coefficient — the weight on an explanatory variable.
intercept — the constant *b* in the equation.
confounder — a hidden third cause behind a correlation.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? Start from the result you want to predict (sales), then ask which measurable things plausibly move it — and for each, ask "could something else be driving both?" A high R^2 tells you the line fits, not that the cause is real.

🌍 In a New Context

A city finds that months with more ice cream sales also have more swimming accidents, and a regression line fits the two closely. Explain why this does not mean ice cream causes accidents, and name the hidden factor most likely behind both.

❓ Lesson Question — Answered

The lesson asked how regression turns data into a prediction and how we judge whether it can be trusted. Regression analysis expresses the relationship between an explanatory variable (the input used to make the prediction) and a response variable (the result) as a model: a regression line $y = ax + b$ in simple regression (one explanatory variable), or a weighted sum such as $y = a_1x_1 + a_2x_2 + b$ in multiple regression (several variables), fitted by least squares to minimize the sum of squared residuals. Trust is judged with evaluation indicators used together — R^2 for how much variability is explained, and RMSE and MAE for the size of prediction errors (RMSE weighting large misses more, MAE less swayed by outliers) — never R^2 alone, and never by adding variables for their own sake. Above all, a good fit is only a correlation: causation requires temporal precedence, ruling out other factors, and theory. Regression offers one perspective from the data, to be interpreted, not obeyed.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

In regression analysis, the variable that is the result is called the (a), and the variable used to predict it is called the (b). The method that predicts using one

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

explanatory variable is called (c), and the method that predicts using multiple explanatory variables is called (d). To judge whether a model is good, evaluation indicators are used. Indicators include (e), which shows how well the model explains the variability in the data, and (f) and (g), which show the size of the discrepancies between predicted and actual values. To claim causation from the results of regression analysis, temporal precedence, exclusion of other factors' influence, theoretical support, and so on are necessary, and one must not conclude that there is (h) just because the prediction equation fits.

1. Fill in the blanks (a)–(h).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes correlation and causation.

A If there is correlation, causation necessarily holds.

B To claim causation, temporal precedence, exclusion of other factors' influence, theoretical support, and so on are necessary.

C If a prediction equation is obtained from regression analysis, the equation completely demonstrates causation.

D If there is no correlation, it can be said that there is no relationship at all between the two variables.

(2) From the following options (A–D), choose the one that most appropriately describes the interpretation of evaluation indicators.

A R^2 takes values between 0 and 1; the closer to 1, the better the variation in the data is explained.

B RMSE and MAE use the same calculation method, and their values always coincide.

C A model with a high R^2 is always a good model in any situation.

D MAE is an indicator that weights large discrepancies more heavily.

(3) From the following options (A–D), choose the one that is NOT an appropriate point of caution when using regression models.

A Do not conclude that there is causation just because the prediction equation fits.

B Confirm accuracy with evaluation indicators and verify the validity of the model in light of the purpose.

C If R^2 is close to 1, one may judge the goodness of the model without checking other indicators.

D Adding more explanatory variables does not necessarily improve accuracy. Appropriate variables must be chosen according to the purpose.

REFLECT & CHALLENGE

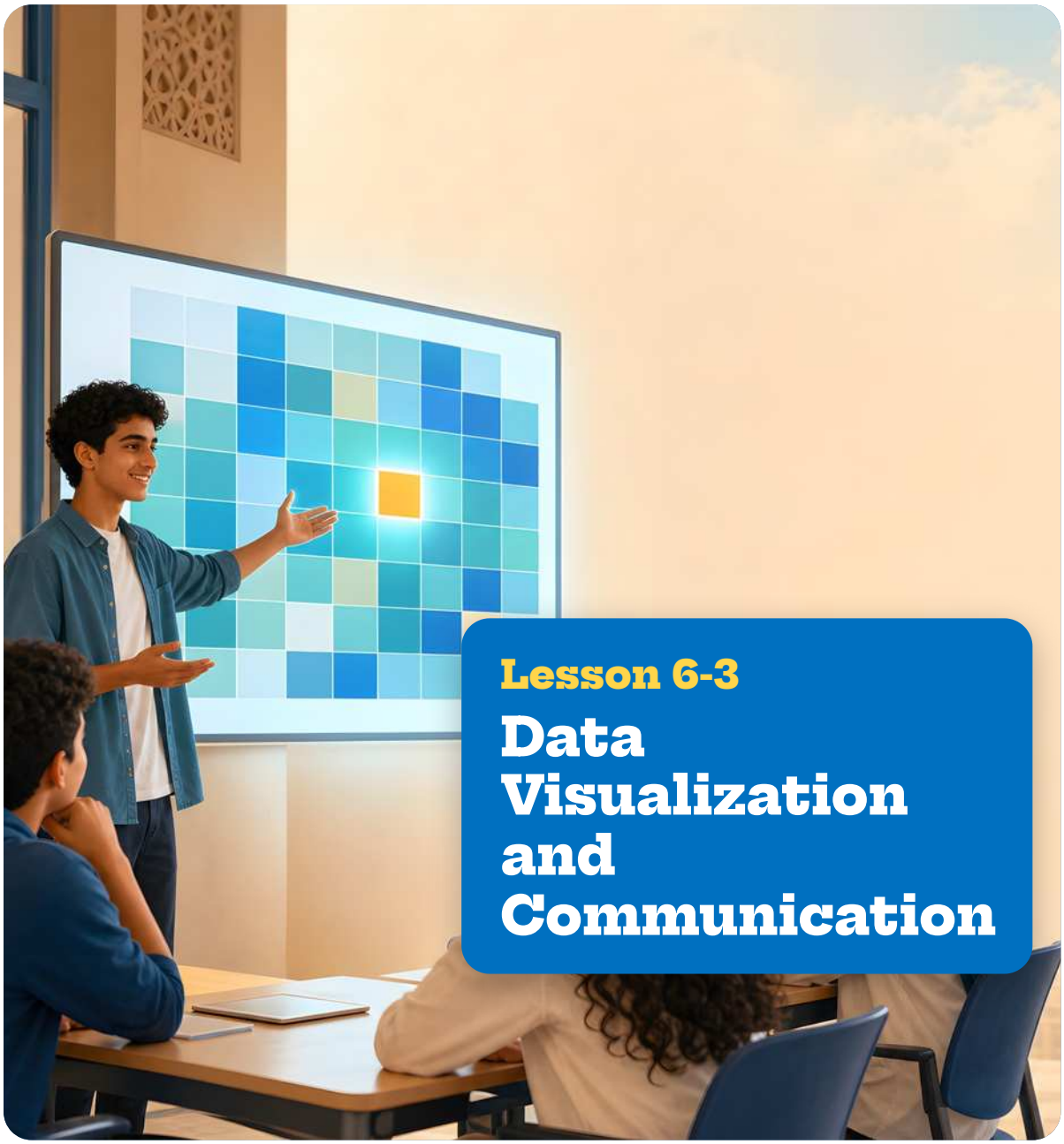
Reflect: Think of two things you have seen rise together (say, screen time and poor sleep). What hidden factor could drive both? **Challenge:** You may add only one more explanatory variable to a model that already has a high R^2. Argue whether you would add it at all, and state the one check you would run before trusting the bigger model.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

⭐ Key takeaway

Remember:
Regression predicts a response variable from one (simple, $y = ax + b$) or several (multiple) explanatory variables, fitted by least-squares. Judge it with R^2, RMSE, and MAE together — and never read a good fit as proof of cause.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 6-3

Data Visualization and Communication

Learning Objectives

- Explain the role of data visualization and select suitable graphs, including for multiple variables.
- Apply data storytelling to communicate a clear and honest message from data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

A table of numbers can hide the very thing you most need to see. Turn it into the right picture and the pattern jumps out — the busiest hours, the strongest link, the odd outlier. Visualization does two jobs: it helps you understand your own data, and it helps you explain it to others. When one or two variables are involved, a bar, line, or scatter plot is enough; when three or more variables, or many combinations of information, must be shown at once, heatmaps and mosaic plots are used instead. And a chart alone is not a message: telling a clear data story is what makes an audience actually understand.

? Guiding question: How do we choose a visualization that reveals a pattern honestly, and how do we turn a chart into a message others understand?

Point!

1. The Role of Data Visualization

(1) **(Data visualization)**…representing data as graphs or figures so that features that are hard to read from numbers alone can be grasped visually.

(2) Data visualization broadly has two roles.

Role	Description	Example
Deepen the analyst's own understanding	Visually grasp the distribution, trends, outliers, etc. of data and gain insight	When training records or study time are shown as a line graph, changes and trends that are hard to notice when numbers are merely listed in a table become easier to see
Communicate data to others	Convey the analysis results or claims to others in an easy-to-understand way	When the average test score of each class is shown as a bar graph, even someone seeing the data for the first time can compare differences between classes at a glance

(3) In either role, it is important to choose a graph that fits the content to be conveyed and to use representations that do not invite misunderstanding. The appropriate visualization method differs depending on the number and type of variables and the purpose of the analysis. When information that cannot be handled by familiar graphs needs to be shown, more advanced visualization techniques are used.

2. Visualization Handling Multiple Variables

Common graphs such as bar graphs, line graphs, and scatter plots are suited to representing the relationship between one or two variables. On the other hand, in analysis, situations often arise in which three or more variables, or many

Key Fact:
Visualization turns numbers into patterns the eye can grasp — to help the analyst think and to communicate. Common graphs suit one or two variables; heatmaps and mosaic plots handle more combinations. Good communication uses data storytelling: a clear message, only the elements that serve it, and honest labels.

Key Concepts:

- data visualization
- heatmap
- mosaic plot
- overplotting
- data storytelling

KEY TERMS
data visualization — showing data as graphs/figures to grasp features visually.

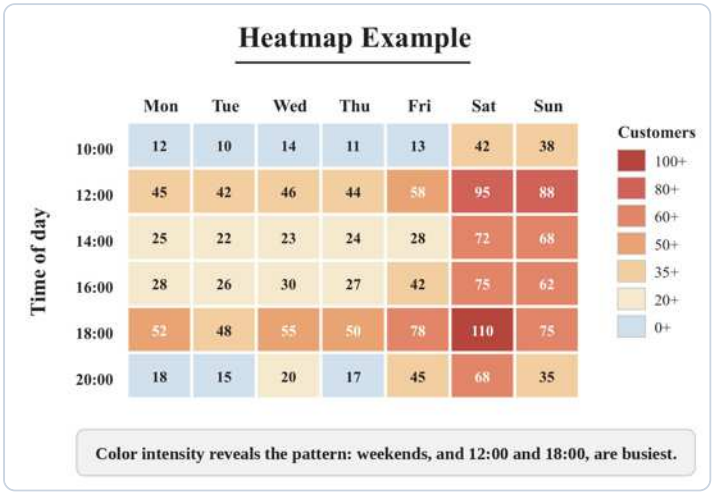
A clear chart speaks; a cluttered one hides.

KEY TERMS
heatmap — a color grid on two axes for a third value.
overplotting — points overlapping so density is hidden.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

combinations of information, need to be handled at once. Heatmaps and mosaic plots are advanced visualization methods used for such cases.

(1) **(Heatmap)**…a graph in which items are arranged on two axes, and the size of a **(third value)** — a numerical value — corresponding to each combination is represented by the gradation of color or differences in hue. *e.g.,* Grasp the number of customers visiting at a glance by day of week × time slot; compare trends in average temperature for city × month.



A heatmap: color intensity shows the value for each row × column combination

In a scatter plot, since the relationship between two quantitative variables is represented by points, when the number of points becomes large, they overlap and the density of data becomes hard to see (overplotting). Also, scatter plots are not suited to handling combinations of categorical variables. In such cases, a heatmap has the advantage of allowing the density and patterns of the data to be grasped at a glance.

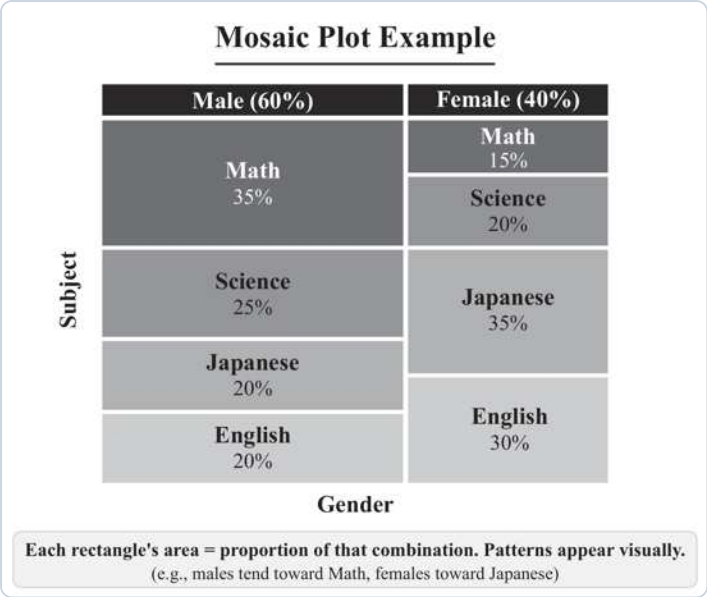
(2) **(Mosaic plot)**…a graph that represents the relationship between two categorical variables by the size of rectangular areas. *e.g.,* Compare the distribution of gender × subject preference; grasp differences in purchase ratios for age group × product category.

One view can encode several variables at once.

PAUSE & THINK

Two categories compared by area, or a third value by color — which graph is which?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



A mosaic plot: each rectangle's area shows the proportion of that category combination

The widths of row and column divisions correspond to the proportions of frequencies, and the area of each rectangle represents the proportion of frequency for that combination of categories. Even when the overall picture is hard to grasp from a numerical table (cross-tabulation), a mosaic plot visually shows differences in proportions as areas, making it easier to grasp relationships between categories intuitively.

3. Data Storytelling

(1) **(Data storytelling)**…using data and visualization to convey the message you want to convey, structured like a story.

(2) The following are the three basic elements of storytelling.

Element	Description	Example
Clarification of the message	Make it possible to express in one sentence what you want to convey through the graph	Make the title of a graph showing class test results contain the claim, such as “The class scored highly overall on the math test”
Selection of elements	Remove elements unrelated to the message and emphasize the parts you want to draw attention to	Make only the elements you want to draw attention to stand out in color, while keeping other elements in faded colors
Provision of supporting information	Properly arrange the title, axis labels, units, and notes	Make the graph title, such as “Average math test scores by class,” express the content at a glance, and clearly indicate units and item names like “points” on the vertical axis and “class” on the horizontal axis

(3) The depth of an audience's understanding can change greatly with how the same analysis result is conveyed. After visualizing data, it is important to put yourself in the audience's position and verify whether the message is conveyed.

KEY TERMS

mosaic plot — rectangle areas for two categorical variables.

data storytelling — structuring a data message like a story.

Storytelling guides the eye to one insight.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Data visualization has both the role of deepening the analyst's own understanding and the role of communicating data to others.

B A heatmap is a graph that represents the size of a third value corresponding to combinations on two axes by the gradation of color or differences in hue.

C A mosaic plot is a graph that represents the relationship between two categorical variables by the size of rectangular areas.

D In data storytelling, it is important to include as many elements related to the graph as possible.

(2) Match each description (a–c) with the most appropriate graph from the choices below (A–C).

a A graph that represents the size of a numerical value corresponding to combinations on two axes by the gradation of color or differences in hue

b A graph that represents the relationship between two categorical variables by the size of rectangular areas

c A graph that represents the relationship between two quantitative variables by the position of points

[Choices] **A** Scatter plot **B** Heatmap **C** Mosaic plot

✓ Solution

(1)

A Correct as a role of data visualization. Therefore, ○.

B Correct definition of a heatmap. Therefore, ○.

C Correct definition of a mosaic plot. Therefore, ○.

D In data storytelling, it is important to remove elements unrelated to the message. Therefore, ×.

(2)

a: Showing values by gradation of color or hue is a heatmap. **B**

b: Showing relationships by area is a mosaic plot. **C**

c: Showing relationships by points is a scatter plot. **A**

PAUSE & THINK

Does adding more elements to a graph make the message clearer, or bury it?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the term for the graph that represents the size of a third value corresponding to combinations on two axes by the gradation of color or differences in hue?
2. What is the term for the graph that represents the relationship between two categorical variables by the size of rectangular areas?
3. In data visualization, of “deepening the analyst's own understanding” and “communicating data to others,” for what purpose is the latter performed?
4. What is the term for using data and visualization to convey the message you want to convey, structured like a story?

2 Answer the following questions.

(1) For each of situations (a, b), choose the most appropriate graph from the choices below (A, B).

a You want to grasp the pattern of customer visits by day of week × time slot at a glance.

b You want to intuitively show differences in proportions for gender × subject preference using areas.

[Choices] **A** Heatmap **B** Mosaic plot

(2) From the following options (A–D), choose the one that most appropriately describes a situation in which a heatmap is used.

A A situation in which one wants to simply compare a representative value for one variable.

B A situation in which the relationship between two quantitative variables can be sufficiently grasped with a scatter plot because the number of points is small.

C A situation in which one wants to grasp the distribution and patterns of numerical data over combinations on two axes at a glance.

D A situation in which one wants to show only one composition ratio of a categorical variable.

EXAM-STYLE QUESTION

Explain why choosing the wrong graph, or hiding a chart's labels, can mislead an audience even when the data is correct. **[6] Refer to:** matching the graph to the data and honest supporting information.

PAUSE & THINK

Two axes with a value by color, or two categories by area? Name each situation's graph.

PAUSE & THINK

Time-series of one value, or proportions of two categories — which suits a mosaic plot?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that most appropriately describes a situation in which a mosaic plot is used.

A A situation in which one wants to follow the time-series changes of one quantitative variable.

B A situation in which one wants to intuitively grasp the proportions of frequencies for combinations of two categorical variables using areas.

C A situation in which one wants to represent the correlation between two quantitative variables by points.

D A situation in which one wants to show the composition ratio of categories at a particular time using sectors.

(4) From the following options (A–D), choose the one that most appropriately describes data storytelling.

A Even if the message you want to convey is unclear, it will be communicated to the audience as long as the graph is beautiful.

B Clarification of the message, selection of elements, and provision of supporting information are the basic elements.

C It is better to include as many related elements as possible in a graph.

D Titles and axis labels detract from the appearance of a graph and should preferably be omitted.

⚙ Think as an Engineer — Investigate & Decide

Your class collected data on students' home governorate and their favorite sport, and also monthly rainfall for several Egyptian cities across the year. You must present each dataset clearly.

☐ **Use a bar graph for everything**

☐ **Match each dataset to the right graph**

1 · Choose a graph for each. For the governorate × favorite-sport data and for the city × month rainfall data, name the graph you would use. For at least one of them, justify your choice — could the other graph type also work here, and why is yours better?

2 · Guard against misleading. Name one design choice that could make your chosen chart mislead the reader (for example a color scale with no legend, or truncated categories), and say how you would prevent it — a judgment the lesson's examples do not hand you.

3 · Tell the story. For one of the two charts, write the one-sentence message you want the audience to take away, and name the single element you would emphasize to carry it. **In pairs,** compare messages and agree on the clearer one.

Give evidence for your answer.

KEY TERMS

cross-tabulation — a table of counts for category combinations.

misleading graph — one whose design distorts the true pattern.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? Ask what each dataset is made of: two categories → areas (mosaic); a value across two axes → color grid (heatmap). Then ask what one sentence you want the reader to leave with, and strip away everything that does not serve it.

🌍 In a New Context

A researcher plots study time against test score for every student in a large school as a scatter plot, but so many points overlap into a solid blob that the density cannot be read. Name a visualization that would reveal where most students actually fall, and explain why it works better here than the crowded scatter plot.

❓ Lesson Question — Answered

The lesson asked how to choose a visualization that reveals a pattern honestly and how to turn a chart into a message others understand. Data visualization serves two roles — deepening the analyst's own understanding and communicating to others — and the right method depends on the number and type of variables. Bar, line, and scatter plots suit one or two variables; when a scatter plot's points overlap (overplotting) or categories must be combined, a heatmap shows a value across two axes by color, and a mosaic plot shows two categorical variables by rectangular area. Turning a chart into a message is data storytelling, built from three elements: clarify the message in one sentence, select only the elements that serve it (emphasizing the key part, fading the rest), and provide honest supporting information (title, axis labels, units, notes). Because the same result can land very differently depending on how it is shown, the final step is to stand in the audience's position and check that the message truly arrives.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Data visualization has both the role of deepening the analyst's own understanding and the role of (a) data to others. When information cannot be

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

handled by common graphs such as bar graphs and line graphs, advanced visualization techniques are used. The graph that represents the size of a third value corresponding to combinations on two axes by the gradation of color or differences in hue is called (b), and the graph that represents the relationship between two categorical variables by the size of rectangular areas is called (c). When communicating data, the concept of (d), in which the message is structured like a story, is important; the basic elements are clarification of the message, selection of elements, and provision of (e).

1. Fill in the blanks (a)–(e).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes a heatmap.

A A graph that represents the relationship between two quantitative variables by the position of points.

B A graph that represents the size of a third value corresponding to combinations on two axes by the gradation of color or differences in hue.

C A graph that represents the composition ratio of categories by the area of sectors.

D A graph that represents the changes in time-series data by lines.

(2) From the following options (A–D), choose the one that is NOT an appropriate basic element of data storytelling.

A Make it possible to express in one sentence what you want to convey through the graph (clarification of the message).

B Remove elements unrelated to the message and emphasize the parts you want to draw attention to (selection of elements).

C Properly arrange the title, axis labels, units, and notes (provision of supporting information).

D Include as many elements unrelated to the message as possible to increase the amount of information.

(3) For each of analysis purposes (a–c), choose the most appropriate graph from the choices below (A–C).

a You want to grasp the monthly average temperatures of multiple cities together at a glance.

b You want to show differences in purchase ratios for age group × product category using areas.

c You want to see the relationship between two quantitative variables in a case where the number of points is small enough that the density can also be sufficiently read.

[Choices] **A** Scatter plot **B** Heatmap **C** Mosaic plot

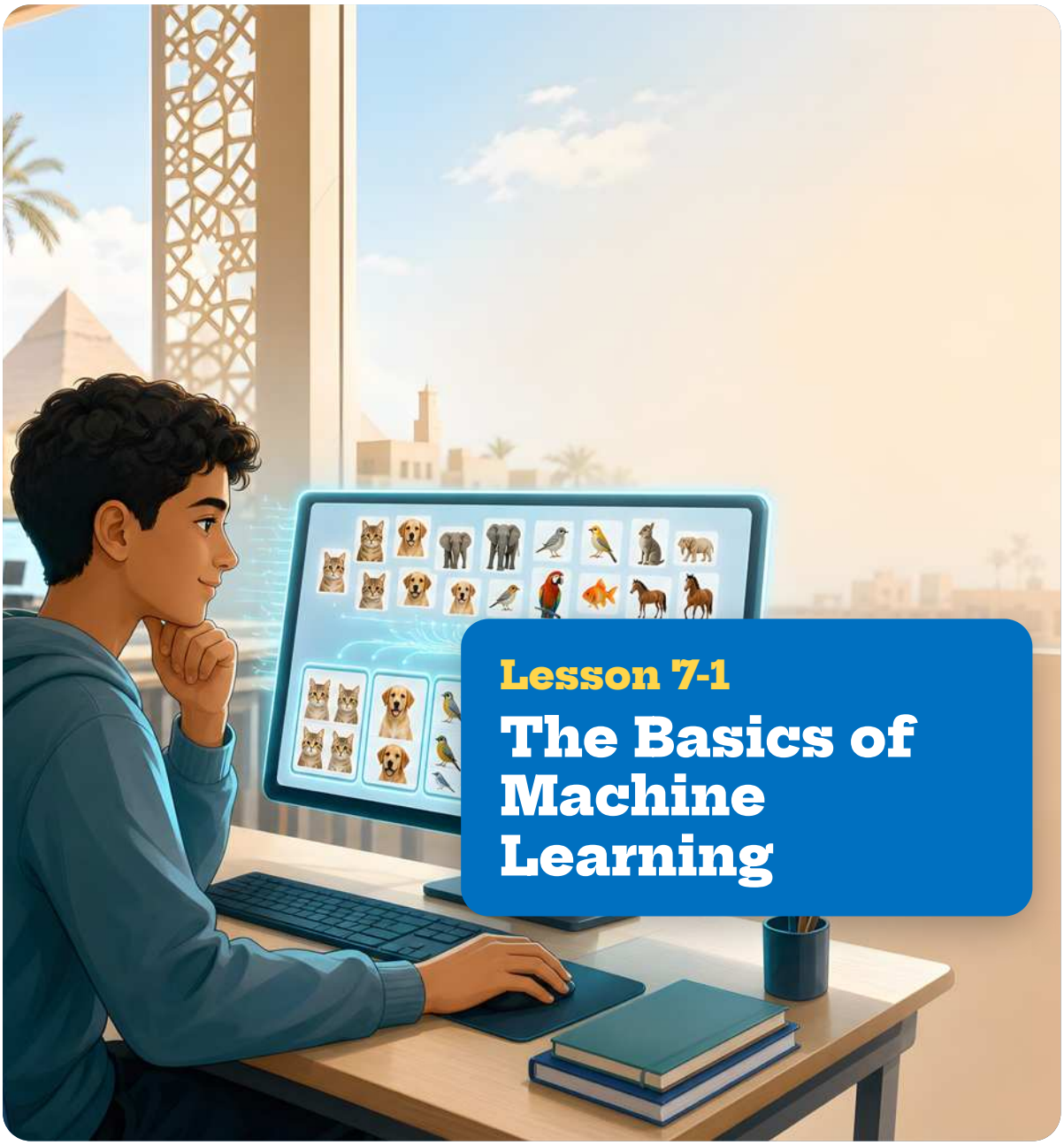
REFLECT & CHALLENGE

Reflect: Recall a chart that confused you. Was it the wrong graph, too many elements, or missing labels? **Challenge:** You are given a heatmap whose color scale does not start at zero, so small differences look dramatic. Describe the false impression this creates, and the one change that would make it honest without changing the data.

Key takeaway

Remember:
Match the graph to the data (heatmap for a value across two axes; mosaic plot for two categories by area; scatter/bar/line for one or two variables), then tell a data story: one message, only the elements that serve it, honest labels — and check it lands with the audience.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 7-1

The Basics of Machine Learning

Learning Objectives

- Explain the idea of machine learning and distinguish supervised, unsupervised, and reinforcement learning.
- Describe classification, clustering, and reinforcement learning with examples, and explain the limits caused by data bias and the black-box problem.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

How do you teach a computer to tell a cat from a dog when you cannot write down a rule for it? You do not write the rule — you show it thousands of examples and let it find the rule itself. That is machine learning, the engine behind most of today's AI. It learns patterns from data to make judgments on new data. Yet it is not magic: if the examples are biased, the judgments are biased, and its reasoning can be a black box even to its makers. Machine learning is therefore best used to support human judgment, not to replace it.

? Guiding question: How does machine learning solve problems differently from ordinary programming, and what are its main types and limits?

Point!

1. The Idea of Machine Learning

(1) **(Machine learning)**…a technology that learns patterns and regularities from large amounts of data so that judgments and predictions can be made on new data. It is the most common method used to make AI work.

(2) Conventional programming and machine learning differ greatly in how they solve problems.

Approach	What humans provide	What the computer does
Conventional programming	Decision rules (procedures) and input data	Produces answers by following the rules
Machine learning	Large amounts of data (with correct answers, where available)	Finds the decision rules from the data on its own

(3) Machine learning is used for problems where it is difficult for humans to express the rules in words. *e.g.,* judging whether a photo shows a dog or a cat; distinguishing whether an email is spam; transcribing audio into text

(4) For such problems, the rules are too complex for humans to write out completely, so the approach of letting the computer learn from examples (data) becomes effective.

(5) However, machine learning is not a panacea. Because the quality and bias of the training data affect the results, biased data leads to biased judgments (bias). There is also the black-box problem, in which the basis for the model's judgments becomes hard for humans to see. Given these limitations, it is important to use machine learning not as a replacement for human judgment, but as a tool that supports human judgment.

2. Three Types of Learning Methods

(1) Machine learning is broadly divided into the following three types, depending on the form of data given and the method of learning.

Key Fact:
Machine learning finds the rules from data instead of being given them. It comes in three forms — supervised (learn from labeled examples), unsupervised (find hidden groups), and reinforcement (learn from rewards) — but its results depend on the data, so bias and the black-box problem make it a support for human judgment, not a replacement.

Key Concepts:

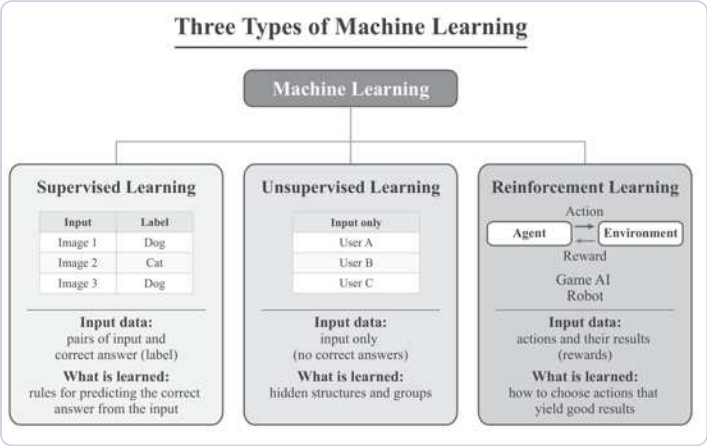
- machine learning
- supervised / unsupervised / reinforcement
- classification, feature
- clustering
- bias, black-box problem

KEY TERMS
machine learning — learning rules from data to judge new data.
bias — skewed data leading to skewed judgments.
black-box problem — the basis of a judgment is hard to see.

Machine learning improves from more examples.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Learning method	Data given	What is learned
(Supervised learning)	Pairs of input and the correct answer (label)	Rules for predicting the correct answer from the input
(Unsupervised learning)	Input only (no correct answers)	Hidden structure or groups in the data
(Reinforcement learning)	Actions and their results (rewards)	How to choose actions to obtain better results



The three types of machine learning and the data each uses

(2) Which learning method to use is determined by the type of data at hand and the nature of the problem to be solved.

Situation	Suitable learning method
Past data has correct answers, and you want to predict the correct answer for new data	Supervised learning
Data is available but without correct answers, and you want to find tendencies or groups in the data	Unsupervised learning
You want the system to learn how to choose actions for better results through trial and error	Reinforcement learning

3. Classification (A Representative Example of Supervised Learning)

(1) **(Classification)**…a method of predicting which type (class) the input data belongs to. It is a representative example of supervised learning. *e.g.,* sorting emails into "spam" or "normal"; sorting medical images into "abnormal" or "normal"; classifying animals in photos as "dog," "cat," or "bird"

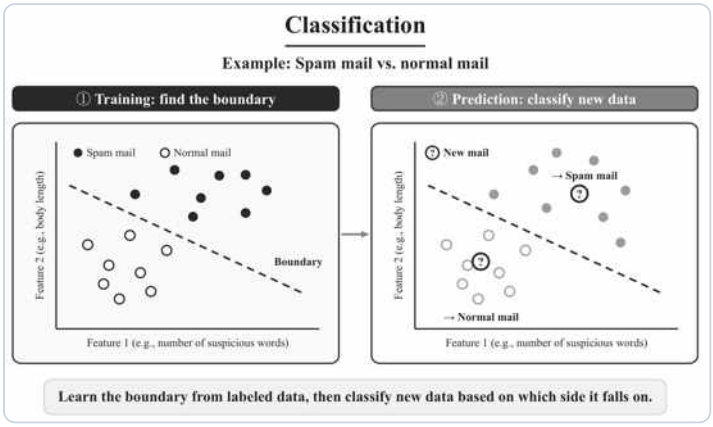
(2) In classification, large amounts of data with correct answers (labels) attached are given in advance for learning. After training is complete, the model can predict which class new data without known labels belongs to.

KEY TERMS
unsupervised learning — finds groups with no labels.
reinforcement learning — learns actions from rewards.
label — the known correct answer on training data.

KEY TERMS
supervised learning — learns from labeled examples.
classification — predict which class an input belongs to.
feature — a clue in the data useful for prediction.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stage	What is done
(Training)	Find rules for classification from pairs of input data and correct answers (labels)
(Prediction)	Use the learned rules to determine which class new data belongs to


Classification
Example: Spam mail vs. normal mail
① Training: find the boundary
② Prediction: classify new data
Spam mail
Normal mail
Boundary
New mail
→ Spam mail
→ Normal mail
Feature 1 (e.g., number of suspicious words)
Feature 2 (e.g., body length)
Learn the boundary from labeled data, then classify new data based on which side it falls on.

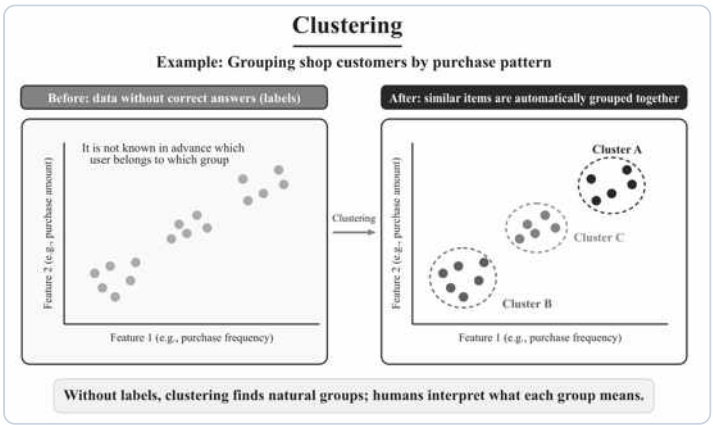
Classification: learn a boundary from labeled data, then classify new data

(3) Information in the training data that serves as a clue useful for prediction is called a **(feature)**. *e.g.,* In spam classification, the types of words contained in the email, the sender's address, and the length of the body are features.

(4) The choice of features greatly affects the accuracy of classification. Choosing features that match the purpose is important for building a good classification model.

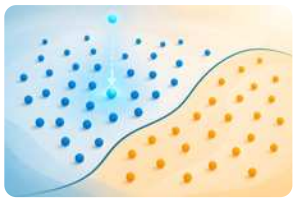
4. Clustering (A Representative Example of Unsupervised Learning)

(1) **(Clustering)**…a method that groups similar items together into groups (clusters) within data. It is a representative example of unsupervised learning. *e.g.,* grouping users of an online shop by similar purchase tendencies; grouping users with similar listening patterns on a music streaming service for recommendations; automatically grouping news articles with similar content


Clustering
Example: Grouping shop customers by purchase pattern
Before: data without correct answers (labels)
After: similar items are automatically grouped together
It is not known in advance which user belongs to which group
Clustering
Cluster A
Cluster C
Cluster B
Feature 1 (e.g., purchase frequency)
Feature 2 (e.g., purchase amount)
Without labels, clustering finds natural groups; humans interpret what each group means.

Clustering: without labels, similar items are automatically grouped

(2) The difference between clustering and classification lies in whether the correct answers (labels) are given.



Classification draws a boundary between labeled groups.

PAUSE & THINK

Labels given or not given — which one word separates classification from clustering?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Method	Correct answers (labels)	Purpose
Classification	Given in advance	Predict which class new data belongs to
Clustering	Not given	Find natural groups within the data

(3) In clustering, the computer itself finds "what is similar." Even when the names or number of groups are not known in advance, grouping can be done automatically based on the features of the data.

(4) Clustering is used when one wants to grasp the overall picture of the data or when one wants to look at the data from a new perspective. Humans interpret the results of grouping and use them in marketing, product development, and so on.

5. Reinforcement Learning

(1) **(Reinforcement learning)**…a method in which actions are chosen in a certain situation, and how to choose better actions is learned based on the reward obtained as a result. *e.g.,* learning how to obtain a high score in a game through trial and error; a robot learning to walk without falling by trying movements; a self-driving car learning safe driving operations

(2) In reinforcement learning, learning is accomplished by repeatedly trying combinations of actions and results.

Element	Description
(State)	The situation in which the agent (the learning side) is placed
(Action)	The choices that can be selected in that situation
(Reward)	The evaluation given as a result of the action (a numerical value indicating goodness or badness)

(3) What distinguishes reinforcement learning from the other two learning types is that it learns from the consequences of its own actions. Unlike supervised learning, where there is a correspondence of "this input has this correct answer," reinforcement learning gradually learns better choices from the rewards obtained for the actions it tries.

(4) Reinforcement learning is suited to problems where the correct answer is not known in advance, but the goodness of the result can be judged. It is often used in combination with environments where trial and error can be repeated (simulators or games).

KEY TERMS

clustering — group similar items with no labels.
reinforcement learning — learn actions from rewards.
reward — a value scoring how good an action was.

Reinforcement learning learns by reward and penalty.

PAUSE & THINK

Learning from rewards through trial and error — which of the three types is that?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A Machine learning is a technology that learns patterns and regularities from large amounts of data so that judgments and predictions can be made on new data.

B In supervised learning, the model learns from data without correct answers (labels).

C Clustering is a method that groups data according to labels given in advance.

D Reinforcement learning is a method in which how to choose better actions is learned based on the reward obtained as a result of an action.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A learning method that learns rules for predicting correct answers for inputs from pairs of input and correct answer

b A method that finds groups of similar items from data without correct answers

c Information in the input data that serves as a clue for prediction in classification

[Choices] **A** Feature **B** Supervised learning **C** Clustering

✓ Solution

(1)

A Correct definition of machine learning. Therefore, ○.

B Data without correct answers (labels) is unsupervised learning. In supervised learning, correct answers are given. Therefore, ×.

C Clustering is a method that finds natural groups from data without given labels. Therefore, ×.

D Correct description of reinforcement learning. Therefore, ○.

(2)

a: Learning from pairs of input and correct answer is supervised learning. **B**

b: Finding groups from unlabeled data is clustering. **C**

c: Information that serves as a clue for prediction is a feature. **A**

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Try

1 Answer the following questions.

1. What is the term for the technology that learns patterns and regularities from large amounts of data so that judgments and predictions can be made on new data?
2. In machine learning, what is the term for the information in the input data that serves as a clue useful for prediction?
3. What is the term for the representative method of unsupervised learning that groups similar items together?
4. In reinforcement learning, what is the term for the numerical value that represents the evaluation given as a result of an action?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that is NOT an appropriate situation in which machine learning is used.

A Distinguishing the type of animal in a photograph.

B Transcribing audio into text.

C Calculating the sum of two integers.

D Distinguishing whether an email is spam.

(2) For each of purposes (a–c), choose the most appropriate learning method from the choices below (A–C).

a Predict, from past purchase history data and whether or not a purchase was made, whether a new customer will buy the product.

b Find groups of customers with similar purchase tendencies from customer purchase data.

c Have a robot walk and learn movements that allow it to advance without falling through trial and error.

[Choices] **A** Unsupervised learning **B** Supervised learning **C** Reinforcement learning

EXAM-STYLE QUESTION

Explain why machine learning should support human judgment rather than replace it. **[6] Refer to:** bias in data and the black-box problem.

PAUSE & THINK

Which task has a simple rule a human can already write — so needs no learning?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that most appropriately describes the difference between classification and clustering.

A Classification divides data into a small number of groups, and clustering divides data into many groups.

B Classification handles only numerical data, and clustering handles only text data.

C Classification learns from data with correct answers (labels), and clustering learns from data without correct answers.

D Classification is unsupervised learning, and clustering is supervised learning.

(4) From the following options (A–D), choose the one that is NOT an appropriate point of caution when using machine learning.

A If the training data is biased, the model's judgments may also be biased.

B The basis for a machine learning model's judgments may be hard to see, so for important judgments human checking is needed.

C Machine learning is a panacea and can make more accurate judgments than humans on any problem.

D It is important to use machine learning as a tool that supports human judgment.

⚙ Think as an Engineer — Investigate & Decide

Your school has years of data and wants help with two tasks: (i) flagging students who may need extra support early, and (ii) discovering natural groups of study habits it did not know about. You advise on how to use machine learning responsibly.

☐ Let the model decide alone **☐ Choose the right method, and keep a human in charge**

1 · Identify the method. For task (i) and task (ii), name which learning type fits (supervised, unsupervised, or reinforcement). For each, name one piece of data the school would need to collect, and one kind of student that data might miss.

2 · Analyze the risks. Name one way the training data for task (i) could be biased (for example, if it only includes students who already asked for help), and say what harm the black-box problem could cause if a teacher acts on a flag without understanding it — a judgment the lesson's tables do not make for you.

3 · Decide the role of the human. State which decisions you would keep a human in charge of, and which you would let the model assist with, and why. **In pairs,** compare plans and agree on the one safeguard you would never remove.

Give evidence for your answer.

KEY TERMS

label — the known correct answer attached to data.

human-in-the-loop — a person checks or approves the model's output.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? Start from the data you have: are there correct answers to learn from (supervised) or none (unsupervised)? Then ask, for each decision, what it would cost to be wrong — the higher that cost, the more a human must stay in charge.

🌍 In a New Context

A bank wants software that reads each loan application and flags the ones that look risky for a human officer to review. Name which learning type this is, name one feature the model might use, and explain why the officer must stay in the loop rather than letting the model decide alone.

❓ Lesson Question — Answered

The lesson asked how machine learning differs from ordinary programming and what its main types and limits are. In conventional programming, humans provide the rules and the computer follows them; in machine learning, humans provide data and examples, and the computer finds the rules itself — which is why it suits problems whose rules are too complex to write, like recognizing images or speech. It comes in three types: supervised learning (labeled examples, whose representative task is classification, using features to predict a class), unsupervised learning (no labels, whose representative task is clustering, finding natural groups), and reinforcement learning (learning from rewards through trial and error, described by state, action, and reward). Its limits matter as much as its power: results depend on the quality and bias of the data, so biased data gives biased judgments, and the black-box problem hides the basis of a decision. For these reasons machine learning is best used to support human judgment, not to replace it.

Exercise

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

Machine learning is a technology that learns patterns and regularities from large amounts of data so that judgments and predictions can be made on new data. Machine learning is broadly divided into three types depending on the form of data given and the method of learning. There is (a) learning, which learns from pairs of input and correct answer; (b) learning, which finds structure or groups from data without correct answers; and (c) learning, which learns how to choose better actions based on the (d) obtained as a result of an action. In (e), a representative example of supervised learning, the choice of (f), which serves as a clue for prediction, greatly affects the result. In (g), a representative example of unsupervised learning, the computer itself finds “what is similar” and discovers natural groups.

1. Fill in the blanks (a)–(g).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the difference between conventional programming and machine learning.

A In conventional programming, humans provide the decision rules; in machine learning, the computer finds the decision rules from data on its own.

B Conventional programming requires large amounts of data, and machine learning requires only rules.

C Conventional programming is included in AI, but machine learning is not included in AI.

D Conventional programming can handle new problems, but machine learning can handle only fixed problems.

(2) For each of situations (a–c), choose the most appropriate term from the choices below (A–C).

a There are large amounts of news article texts and their categories (politics, economics, sports, etc.), and you want to automatically determine the category of new articles.

b You have records of products purchased by users of an online shop and want to find groups of users with similar purchasing tendencies.

c You want a computer to play a game and learn the operations that lead to high scores through trial and error.

[Choices] **A** Reinforcement learning **B** Clustering **C** Classification

REFLECT & CHALLENGE

Reflect: Where have you likely met machine learning today (recommendations, photo tagging, voice input)? Which type do you think it was?
Challenge: A face-recognition model was trained mostly on one group of people and works poorly on others. Explain how biased training data caused this, and state one change to the data that would reduce the bias.

Key takeaway

Remember:
Machine learning finds rules from data: supervised (labeled → classification with features), unsupervised (unlabeled → clustering), reinforcement (rewards). Mind its limits — bias and the black-box problem — and keep it supporting human judgment.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that is NOT an appropriate description of the characteristics and limitations of machine learning.

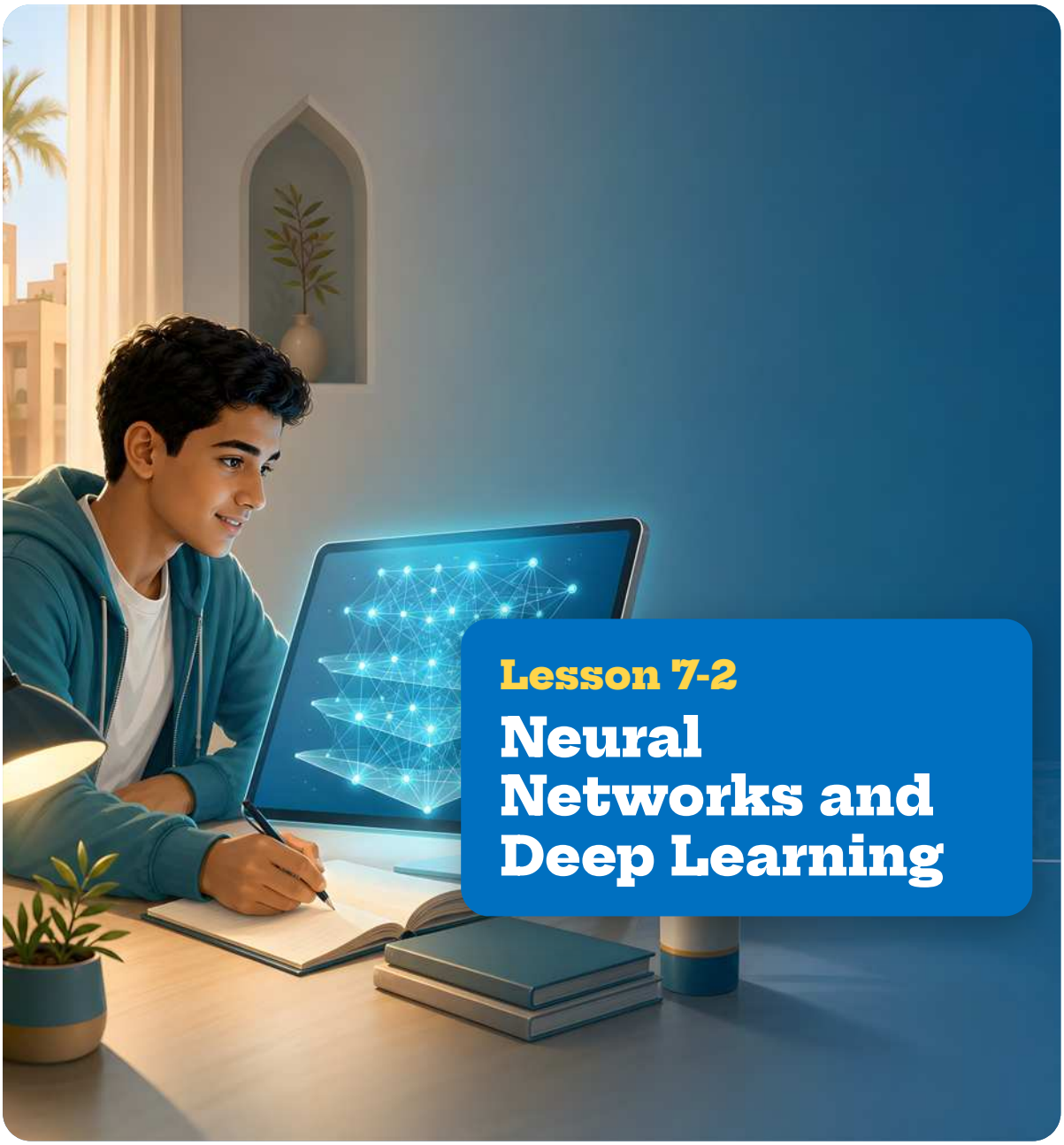
A Machine learning is used for problems where it is difficult for humans to express the rules in words.

B Machine learning models are affected in their results by the quality and bias of the training data.

C A machine learning model's basis for judgment is always shown to humans in an easy-to-understand way.

D It is important to use machine learning not as a replacement for human judgment, but as a tool that supports human judgment.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 7-2

Neural Networks and Deep Learning

Learning Objectives

- Describe the structure of a neural network and explain how a single neuron processes its inputs.
- Explain what deep learning is, and why stacking many hidden layers brought a leap in AI performance.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

How can a machine tell a dog from a cat, or turn speech into text? The answer, loosely inspired by the brain, is a neural network: layers of simple units called neurons, each passing signals to the next. A neuron just multiplies its inputs by weights and adds them up — nothing complex on its own — but stack enough of them across enough layers and the network learns, by adjusting those weights on data, to recognize patterns no human could write rules for. Making the stack deep — deep learning — lets the network build up features from simple edges up to whole objects, and it is what pushed AI past humans on many tasks.

? Guiding question: How is a neural network built and trained, and what does making it "deep" add?

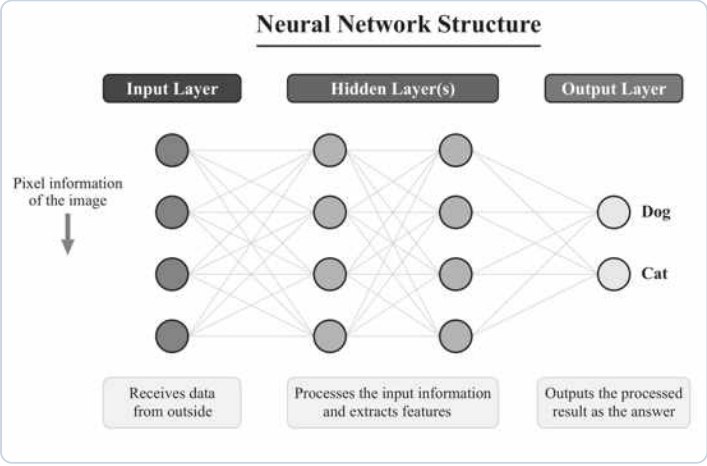
Point!

1. The Structure of Neural Networks

(1) A neural network has a structure in which many **(nodes)**, also called neurons, are connected. Nodes are arranged together in **(layers)**, and the nodes of adjacent layers are connected to each other.

(2) Layers are divided by role into the following three types.

Type of layer	Role
Input layer	Receives data from outside
Hidden layer (middle layer)	Processes the input information and extracts features
Output layer	Outputs the result of processing as an answer



A neural network: input layer, hidden layers, and output layer

(3) Data enters at the input layer, is processed as it passes through the hidden layers, and reaches the output layer in this flow. *e.g.,* In a network that judges whether the animal in a photograph is a dog or a cat, the input layer receives the

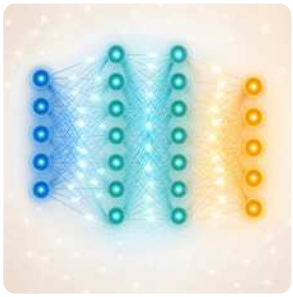
Key Fact:
A neural network is layers of connected neurons: an input layer, hidden layers that extract features, and an output layer. Each neuron multiplies its inputs by weights and sums them; training adjusts those weights from data. Deep learning stacks many hidden layers to build features step by step — the leap behind modern AI.

Key Concepts:

- node (neuron), layer
- input / hidden / output layer
- weight, training
- deep learning
- hierarchical features

KEY TERMS

node (neuron) — a single processing unit.
layer — a group of neurons at one stage.
input / hidden / output layer — receive, process, answer.



A network stacks neurons into connected layers.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

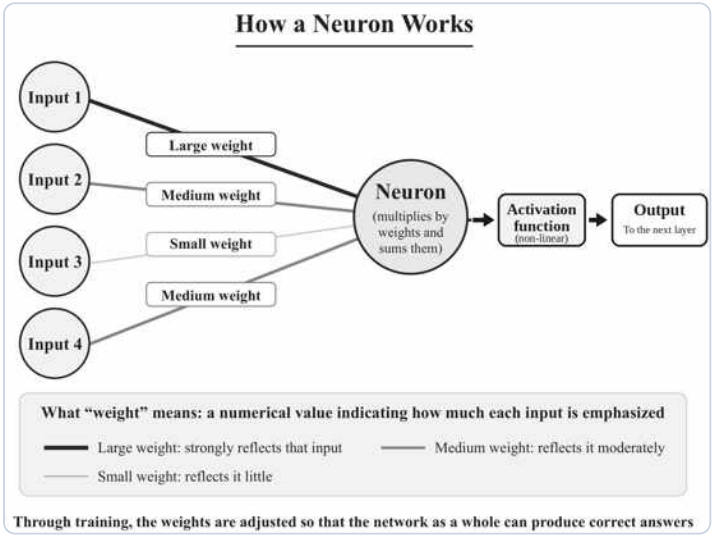
pixel information of the photograph, the hidden layers extract features such as shape and pattern, and the output layer outputs the judgment of “dog” or “cat.”

(4) The number of hidden layers and the number of nodes in each layer are designed according to the difficulty of the problem to be handled. The more complex the problem, the more layers and nodes are needed.

2. How a Neuron Works

(1) Each neuron receives multiple inputs, multiplies them by weights and sums them, then passes the sum through an **(activation function)**, which lets the network learn relationships that are not simply proportional, before passing the result to the next layer.

(2) **(Weight)**…a numerical value that represents how much each input is emphasized. Weights can be negative as well as positive, and the larger a weight is in size, the more strongly that input influences the neuron's result.


How a Neuron Works
Input 1
Input 2
Input 3
Input 4
Large weight
Medium weight
Small weight
Medium weight
Neuron
(multiplies by weights and sums them)
Activation function
(non-linear)
Output
To the next layer
What “weight” means: a numerical value indicating how much each input is emphasized
Large weight: strongly reflects that input
Medium weight: reflects it moderately
Small weight: reflects it little
Through training, the weights are adjusted so that the network as a whole can produce correct answers

A neuron weights its inputs, sums them, applies an activation function, and passes the result on

(3) **(Training)** is the process of gradually adjusting the weights between neurons using a large amount of data.

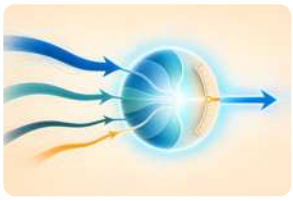
Stage	What happens
Before training	The weights are set to arbitrary values, and correct predictions cannot be made
During training	Data is given, and the weights are gradually adjusted based on the difference between the prediction and the correct answer
After training	With the adjusted weights, correct predictions can be made even on new data

(4) As training progresses, a combination of weights that allows the network as a whole to produce correct answers is found. Even though the function of each individual neuron is simple, complex judgments become possible by combining many neurons.

KEY TERMS

weight — how much an input is emphasized.

training — adjusting weights from data.



A neuron weights its inputs, sums them, applies an activation function, and passes the result on.

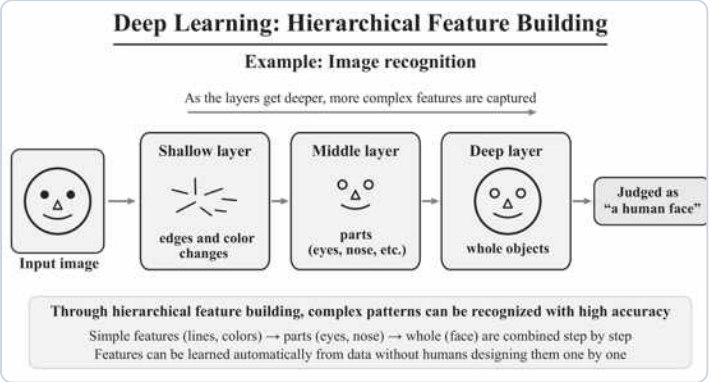
D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

3. Deep Learning

(1) **(Deep learning)** is a technology that uses neural networks with many hidden layers stacked deeply. “Deep” refers to having many hidden layers.

(2) When the hidden layers are made deeper, features of data can be built up step by step and captured. *e.g.,* In an image-recognition network, features are built up from shallow to deep layers as follows.

Layer position	Examples of features extracted
Shallow layers	Simple features such as lines and changes in color
Middle layers	Shapes of parts such as eyes, nose, and ears
Deep layers	Structure of the whole face or object


Deep Learning: Hierarchical Feature Building
Example: Image recognition
As the layers get deeper, more complex features are captured
Shallow layer
edges and color changes
Middle layer
parts (eyes, nose, etc.)
Deep layer
whole objects
Judged as “a human face”
Input image
Through hierarchical feature building, complex patterns can be recognized with high accuracy
Simple features (lines, colors) → parts (eyes, nose) → whole (face) are combined step by step
Features can be learned automatically from data without humans designing them one by one

Deep learning builds features step by step, from edges to whole objects

(3) Through this hierarchical building of features, deep learning can recognize complex patterns with high accuracy.

(4) Deep learning became practical as a result of large amounts of data and high computational power (in particular, specialized processors well suited to AI computation, such as GPUs and TPUs) becoming available. Before that, it was technically difficult to advance learning sufficiently even when layers were made deeper.

4. Why Deep Learning Brought a Leap in AI

(1) With the availability of deep learning, accuracy close to or surpassing humans has become achievable in many tasks that were previously considered difficult for computers.

(2) The fields in which there has been particularly great progress include the following.

PAUSE & THINK

Why does stacking more hidden layers let a network capture more complex features?

Deep layers build from edges to whole objects.

PAUSE & THINK

What two things had to become available before deep learning could work in practice?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Field	Examples
(Image recognition)	Face recognition systems, diagnostic support for medical images, object detection for self-driving cars
(Speech recognition)	Voice operation of smart speakers, transcription from speech, speech translation
(Natural language processing)	Machine translation, summarization of text, answering questions

(3) Before deep learning, features had to be designed by humans one by one and given to the computer. With deep learning, features can be learned automatically from data, allowing the limitations of human design to be exceeded.

(4) As a result of accumulating these capabilities, the foundation has been laid for AI that newly creates text and images.

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A neural network has three types of layers: input layer, hidden layer, and output layer.

B The input layer is the layer that outputs the results of processing as an answer to the outside.

C Deep learning is a technology that uses a neural network with only one hidden layer.

D Training a neural network is the process of adjusting the weights between neurons using a large amount of data.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a In a neural network, the layer that receives data from outside

b In a neural network, a numerical value that represents how much each input is emphasized

c In a neural network, the layer that processes input information and extracts features

[Choices] **A** Hidden layer **B** Weight **C** Input layer

✓ Solution

(1)

A There are three types of layers: input, hidden, and output. Therefore, ○.

B The output layer is what outputs results to the outside. The input layer is the layer that receives data from outside. Therefore, ×.

C Deep learning is a technology that uses neural networks with many hidden layers stacked deeply. Therefore, ×.

PAUSE & THINK

Which layer receives the data, and which one gives the final answer?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

D Correct description of training. Therefore, ○.

(2)

a: The one that receives data is the input layer. **C**

b: The value representing how much an input is emphasized is the weight. **B**

c: The one that processes information and extracts features is the hidden layer. **A**

Try

1 Answer the following questions.

1. What is the term for the individual processing unit that makes up a neural network?
2. In a neural network, what is the term for the layer that processes input information and extracts features?
3. In a neural network, what is the term for the numerical value that represents how much each input is emphasized?
4. What is the term for the technology that uses neural networks with many hidden layers stacked deeply?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the layers of a neural network.

A The input layer has the role of outputting processing results to the outside.

B The hidden layer has the role of receiving data from outside.

C The output layer has the role of processing input information and extracting features.

D Data enters at the input layer, is processed as it passes through the hidden layers, and reaches the output layer in this flow.

(2) From the following options (A–D), choose the one that most appropriately describes the training of a neural network.

A From before training begins, the weights between neurons are set to correct values.

B As training progresses, a combination of weights that allows the network as a whole to produce correct answers is found.

C Training is performed by increasing the number of neurons.

D Once training is complete, the network becomes unable to handle new data.

EXAM-STYLE QUESTION

Explain how a neural network with simple neurons can make complex judgments, referring to weights and training. **[6]**

Refer to: weighting inputs and adjusting weights on data.

PAUSE & THINK

Before training, are the weights already correct? What changes them?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that is NOT an appropriate characteristic of deep learning.

A It has a structure with many hidden layers stacked deeply.

B It can build up features of data step by step.

C The deeper the hidden layers, the more complex features can be captured.

D Even if the number of hidden layers is increased, the difficulty of problems that can be handled does not change.

(4) For a deep learning network that performs image recognition, reorder the following (A–C) in the order in which features are extracted from shallow layers to deep layers.

A Structure of the whole face or object

B Simple features such as lines and changes in color

C Shapes of parts such as eyes, nose, and ears

PAUSE & THINK

More hidden layers let a network capture more complex features — but what does that extra depth demand in data and computing power?

Think as an Engineer — Investigate & Decide

A team wants an app that reads handwritten digits 0–9 from scanned school registers and turns them into typed numbers. You advise on building it with a deep learning network.

☐ **Write the rules for each digit by hand**

☐ **Train a deep network on many examples**

1 · Design the layers. Say what the input layer would receive, what the output layer would produce, and in one line why hidden layers (rather than a single step) are needed here.

2 · Explain the depth. Using the idea of hierarchical features, describe what shallow, middle, and deep layers might each capture for a handwritten digit — a chain the lesson shows for faces but does not spell out for digits, so you must map it yourself.

3 · Judge the cost and risk. Name the two things this network needs to become practical, and name one way the training data could be too narrow, making the app fail on some students' handwriting. **In pairs,** compare designs and agree on the one dataset change that would most improve fairness.

Give evidence for your answer.

Stuck? Use the lesson's face example as a model, then translate it to a digit: ask what would count as an “edge,” a “part,” and a “whole” for a written number. For the risk, list whose handwriting the network saw during training — then ask whose it never saw.

KEY TERMS

training data — the labeled examples a network learns from.

generalize — work correctly on new, unseen data.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

🌍 In a New Context

A voice assistant turns spoken commands into text and often works well even for words it was not explicitly programmed with. Name which field of AI progress this belongs to, and explain, using layers and training, how it can handle speech no one wrote rules for.

❓ Lesson Question — Answered

The lesson asked how a neural network is built and trained and what making it deep adds. A neural network is built from layers of connected neurons: an input layer that receives data, hidden layers that process it and extract features, and an output layer that gives the answer, with data flowing input → hidden → output. Each neuron is simple — it multiplies its inputs by weights, sums them, and passes the sum through an activation function — and the network is trained by gradually adjusting those weights on large amounts of data. Before training, arbitrary weights give wrong answers; after training, tuned weights give correct ones even on new data. Making the network deep — deep learning, many hidden layers — lets features be built hierarchically, from simple edges through parts to whole objects, so complex patterns are recognized with high accuracy. Once large data and powerful processors like GPUs made it practical, deep learning let features be learned automatically instead of designed by hand, driving leaps in image recognition, speech recognition, and natural language processing.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

A neural network has a structure in which many processing units called (a) are connected. These are arranged together in (b), and are divided into three types: (c), which receives data from outside; (d), which processes information and extracts features; and (e), which outputs the result as an answer. Each (a) multiplies its inputs by (f), sums them, passes the sum through an activation function, and passes the result to the next layer. Training gradually adjusts (f) using large amounts of data. The technology that uses neural networks with many hidden layers stacked deeply is called (g), and its characteristic is that it can build up features of data (h) and capture them.

⭐ REFLECT & CHALLENGE

Reflect: Where have you used a tool built on deep learning (voice typing, photo search, translation)? Which field was it?
Challenge: A team makes a network deeper and deeper but accuracy stops improving and training becomes very slow. Using what made deep learning practical, explain what their deeper network now demands more of, and name the one thing — besides adding layers — they should check first.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

1. Fill in the blanks (a)–(h).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes how a neuron in a neural network works.

A Each neuron passes one input as is to the next layer.

B Each neuron multiplies its inputs by weights, sums them, passes the sum through an activation function, and passes the result on.

C Each neuron averages the input values and passes the result to the next layer.

D Each neuron passes only the largest input to the next layer.

(2) From the following options (A–D), choose the one that most appropriately describes the background to deep learning becoming practical.

A Computation became simpler by reducing the number of hidden layers to one.

B Large amounts of data and high computational power became available.

C Connections between neurons were eliminated, simplifying processing.

D A method in which humans set weights one by one by hand was established.

(3) From the following options (A–D), choose the one that most appropriately describes why deep learning brought a leap in AI.

A Features could be learned automatically from data without humans designing them one by one.

B Computers became able to think with the same mechanism as the human brain.

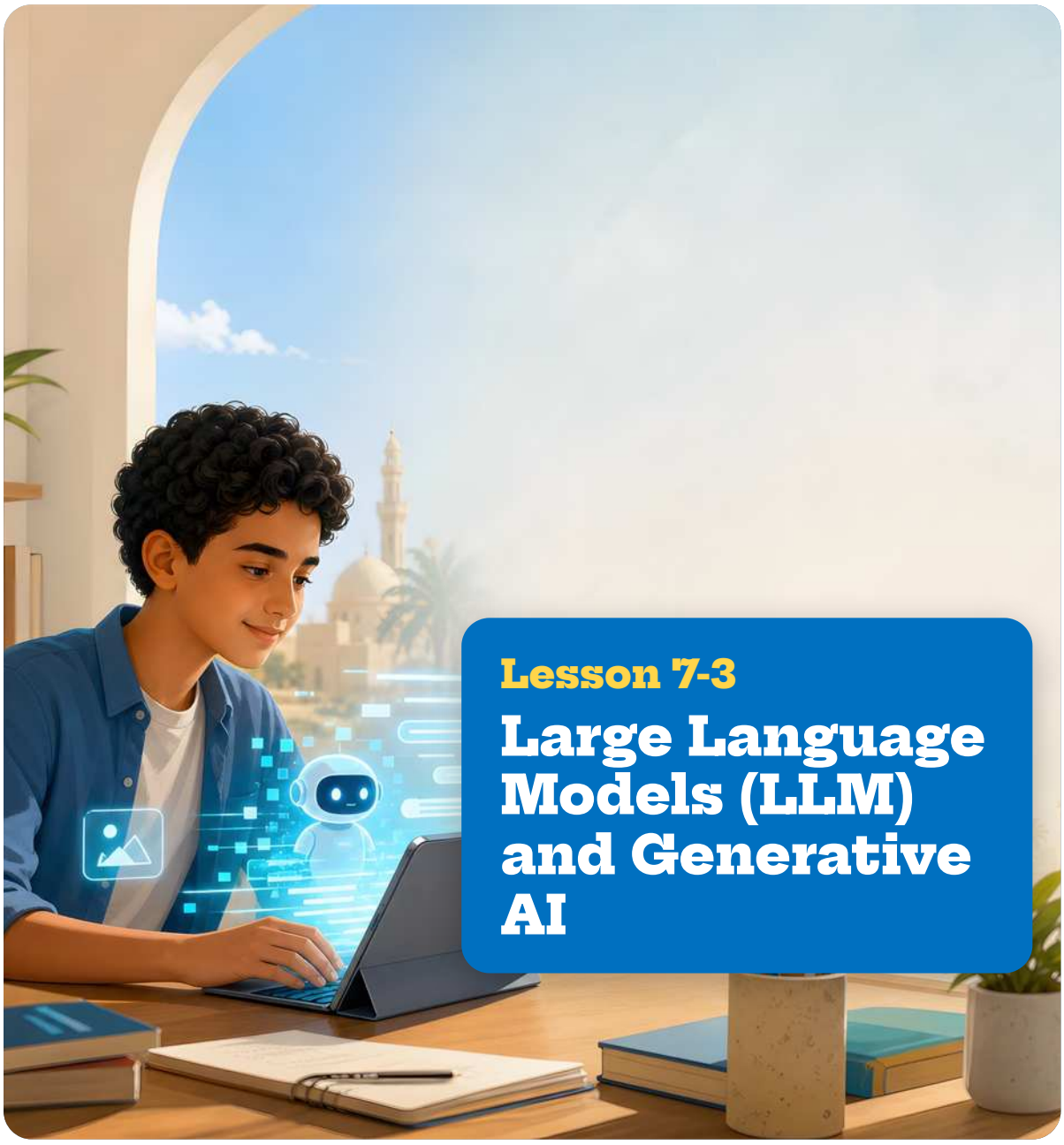
C Computers became able to fully understand the meaning of words.

D Computers became able to make completely accurate judgments on any problem.

⭐ Key takeaway

Remember:
A neural network = layers of neurons (input → hidden → output); each neuron weights and sums its inputs, and training tunes the weights on data. Deep learning stacks many hidden layers to build features (edges → parts → wholes), enabled by big data and GPUs.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي



Lesson 7-3

Large Language Models (LLM) and Generative AI

Learning Objectives

- Explain how a language model generates text by predicting the next word, and why this causes hallucination.
- Describe what makes a language model large, and explain the roles of RLHF tuning and multimodality.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

When a chatbot writes a fluent paragraph, it is not looking up an answer — it is guessing the next word, then the next, using patterns learned from enormous amounts of text. That simple trick, scaled up to billions of parameters, gives large language models their power to translate, summarize, and converse. But because each word is chosen for being statistically plausible rather than certainly true, these models can state falsehoods convincingly — a failure called hallucination. This lesson shows how such models work, how human feedback (RLHF) makes them helpful, and how multimodality lets them handle images and audio too.

? Guiding question: How does a language model generate text, what makes a model "large," and why can it still get facts wrong?

Point!

1. Language Models and "Predicting the Next Word"

(1) **(Language model)**…a model created to learn from large amounts of text and predict which word is likely to come next in a sentence. Through training, it acquires patterns of word usage and phrasing according to context.

(2) Text generation is achieved by repeating the prediction of the next word.

Step	Input sentence	Word predicted next
1st	"Today's weather is"	"sunny"
2nd	"Today's weather is sunny"	"and"
3rd	"Today's weather is sunny and"	"warm"

In this way, a complete sentence is built up by connecting the results of predicting one word at a time.

Key Fact:
A language model generates text by predicting the next word again and again — choosing what is statistically plausible, not guaranteed true (the cause of hallucination). Large language models scale this up with vast data and billions to trillions of parameters, are tuned by RLHF to be helpful, and, through multimodality, handle text, images, and audio together.

Key Concepts:

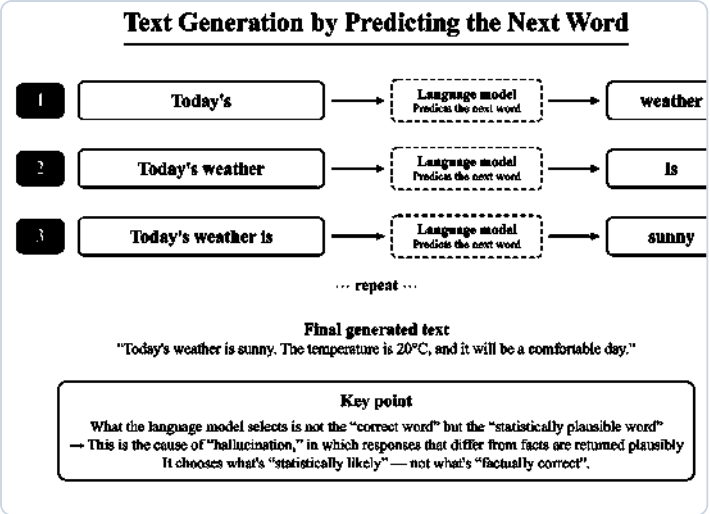
- language model, next-word prediction
- hallucination
- large language model (LLM), parameters
- RLHF
- multimodality

KEY TERMS

language model — predicts the likely next word.
hallucination — a plausible answer that is factually wrong.

A language model predicts the next word.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي


Text Generation by Predicting the Next Word
1
Today's
Language model
Predicts the next word
weather
2
Today's weather
Language model
Predicts the next word
is
3
Today's weather is
Language model
Predicts the next word
sunny
··· repeat ···
Final generated text
"Today's weather is sunny. The temperature is 20°C, and it will be a comfortable day."
Key point
What the language model selects is not the "correct word" but the "statistically plausible word"
→ This is the cause of "hallucination," in which responses that differ from facts are returned plausibly
It chooses what's "statistically likely" — not what's "factually correct".

Text is generated by predicting the next word, one at a time

(3) For each candidate word, the language model calculates a "probability of coming next" and selects words based on that probability. These probabilities come from patterns the model learned during training, so it can also handle wordings that never appeared in the training data.

(4) What a language model selects is not the correct word, but a statistically plausible word. Because of this property, **(hallucination)**—the phenomenon of returning responses that differ from the facts but are stated in a plausible way—occurs.

2. Large Language Models (LLM)

(1) **(Large language model (LLM))**…a language model based on a very large neural network trained using vast amounts of text from the Internet.

(2) "Large" mainly refers to the following two kinds of size.

Aspect	Description
Amount of training data	The model is trained using vast amounts of text containing billions to trillions of words
Number of parameters	The numerical values corresponding to the weights of the neural network number in the billions to trillions

(3) Researchers have observed that when the number of parameters or the amount of training data exceeds a certain scale, qualitatively new abilities begin to appear. *e.g.,* handling complex reasoning; summarizing text; translating into foreign languages; generating program code

(4) Major services built on large language models include ChatGPT, Claude, and Gemini, which provide natural conversation and a wide range of text generation.

3. Tuning by RLHF

(1) Just by learning from vast amounts of text, large language models do not necessarily return responses that are desirable for humans. They may answer

KEY TERMS

large language model (LLM) — a huge network trained on vast text.

parameter — a weight; LLMs have billions to trillions.

RLHF — tuning by human feedback as reward.

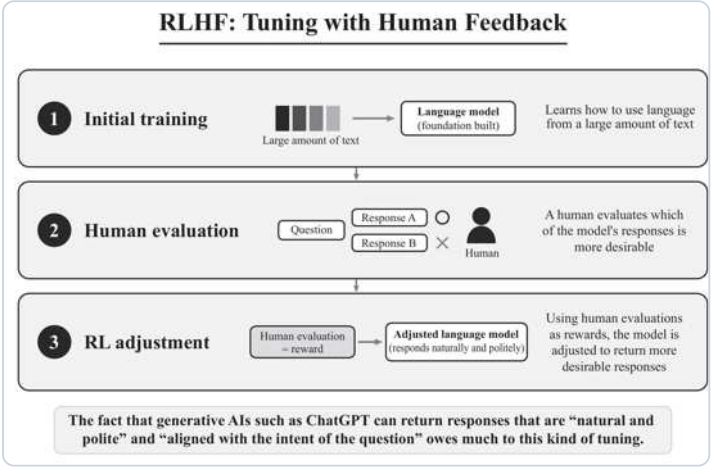
D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

without grasping the intent of the question or return responses containing inappropriate content.

(2) A representative method for solving such problems is **(RLHF (Reinforcement Learning from Human Feedback))**. RLHF is a method that applies reinforcement learning to the tuning of language models.

(3) RLHF proceeds in the following flow.

Stage	Description
① **Initial training**	Build a language model that has learned word usage from large amounts of text
② **Human evaluation**	Humans give evaluations on the model's responses, such as "which is more desirable"
③ **Tuning by reinforcement learning**	Using human evaluations as rewards, tune the model to return more desirable responses


RLHF: Tuning with Human Feedback
1 Initial training
Large amount of text
Language model (foundation built)
Learns how to use language from a large amount of text
2 Human evaluation
Question
Response A
Response B
Human
A human evaluates which of the model's responses is more desirable
3 RL adjustment
Human evaluation = reward
Adjusted language model (responds naturally and politely)
Using human evaluations as rewards, the model is adjusted to return more desirable responses
The fact that generative AIs such as ChatGPT can return responses that are "natural and polite" and "aligned with the intent of the question" owes much to this kind of tuning.

RLHF: human evaluations become rewards that tune the model

(4) Through RLHF, language models begin to return natural and polite responses, responses aligned with the intent of the question, and responses that avoid inappropriate content. The fact that generative AIs such as ChatGPT feel "easy to use" to users owes much to such tuning.

4. Multimodality

(1) **(Multimodality)**…the property of being able to handle multiple kinds of data with a single model, such as text, images, audio, and video. Each kind of data is called a **(modality)**.

(2) Early language models handled only text, but recent generative AIs can handle different kinds of data in combination. *e.g.,* looking at an image and describing its content in text; listening to audio and transcribing it into text; giving instructions in text and generating an image; summarizing the content of a video

(3) The background to multimodality being achieved is that neural networks have become able to handle different kinds of data by converting them into a

PAUSE & THINK

Left alone, a huge model may answer rudely or off-point. What step fixes that?

Human feedback ranks outputs to tune the model.

KEY TERMS

multimodality — one model handling many data kinds.
modality — one kind of data (text, image, audio…).

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

common numerical representation. Inside the network, text, images, and audio are all processed in the same way as sets of numerical values.

(4) Through multimodality, generative AI is evolving from a technology that handles only language into a technology that integrally handles the various kinds of information that humans deal with daily.

One model can output text, image, and sound.

PAUSE & THINK

If a model sounds confident and fluent, does that guarantee it is correct?

✎ Worked Example

(1) Mark each statement (A–D) with ○ if true or × if false.

A A language model is a model that predicts which word is likely to come next in a sentence.

B A language model finishes processing after predicting just one word and does not generate text.

C In a large language model, when the number of parameters or the amount of training data exceeds a certain scale, qualitatively new abilities sometimes appear.

D Multimodality is the property of handling only text.

(2) Match each description (a–c) with the most appropriate term from the choices below (A–C).

a A language model based on a very large neural network trained using vast amounts of text from the Internet

b A method that uses human feedback as a reward to tune a language model to return more desirable responses

c The phenomenon in which a language model returns responses that differ from the facts in a plausible way

[Choices] **A** Hallucination **B** Large language model **C** RLHF

✓ Solution

(1)

A Correct definition of a language model. Therefore, ○.

B Text generation is achieved by repeating the prediction of the next word. Therefore, ×.

C Correct description of large language models. Therefore, ○.

D Multimodality is the property of being able to handle multiple kinds of data—text, images, audio, video, etc.—with a single model. Therefore, ×.

(2)

a: A huge language model trained on vast text is a large language model. **B**

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

b: Tuning by reinforcement learning using human feedback as a reward is RLHF. **C**

c: The phenomenon of returning fact-discrepant responses in a plausible way is hallucination. **A**

Try

1 Answer the following questions.

1. What is the term for the model that learns from large amounts of text and predicts which word is likely to come next in a sentence?
2. What is the term for a language model based on a very large neural network trained using vast amounts of text from the Internet?
3. What is the term for the method that uses human feedback as a reward to tune a language model to return more desirable responses?
4. What is the term for the property of being able to handle multiple kinds of data—text, images, audio, video, and so on—with a single model?

2 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes how a language model works.

A It memorizes the input sentence as is and returns the same sentence.

B It generates text by repeatedly predicting the next word in the sentence.

C It selects and returns the closest match from prepared template sentences.

D It fully understands the meaning of the sentence and returns the single correct response.

(2) From the following options (A–D), choose the one that most appropriately describes the cause of hallucination.

A Because the language model always retrieves and returns sentences from the training data as is.

B Because the language model selects and returns words that are statistically plausible rather than correct.

C Because the language model avoids answering questions and returns unrelated text.

D Because the language model is built to return only fact-based responses.

EXAM-STYLE QUESTION

Explain why a language model can produce a fluent answer that is factually wrong (hallucination). **[6] Refer to:** choosing statistically plausible words.

PAUSE & THINK

"Plausible" or "correct" — which one does the model actually optimize for?

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

(3) From the following options (A–D), choose the one that most appropriately describes RLHF.

A A method of building a language model from scratch using a large amount of text.

B A method of tuning a language model to return more desirable responses, using human evaluations as rewards.

C A method that removes the neural network structure from a language model and simplifies the processing.

D A method that has a language model solve math problems to improve its computational accuracy.

(4) From the following options (A–D), choose the one that is NOT an appropriate application of multimodal AI.

A Showing an image and having it described in text.

B Listening to audio and transcribing it into text.

C Giving instructions in text and generating an image.

D Computing the result of input numerical values using only the rules of the four arithmetic operations and returning it.

⚙ Think as an Engineer — Investigate & Decide

A school club wants to use a generative-AI chatbot to help write short articles about local history for the school website. You advise on using it responsibly.

☐ **Publish whatever the AI writes** ☐ **Use the AI, but verify before publishing**

1 · Decide where it helps. Name one part of the task where the chatbot genuinely saves effort, and one part where its output must not be trusted as-is. Explain each in one line.

2 · Analyze the risk. Using how a language model chooses words, explain why the chatbot might state a wrong date or invented name for a local landmark very confidently, and describe exactly how the club would catch such a hallucination before publishing — a check the lesson does not spell out for you.

3 · Decide the safeguards. State the rule the club should follow for every AI-written article before it goes live, and say what role RLHF-style tuning does and does not solve here. **In pairs,** compare rules and agree on the one check you would never skip.

Give evidence for your answer.

KEY TERMS

generative AI — AI that creates new text, images, or audio.

fact-check — verifying a claim against a reliable source.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Stuck? Remember the model is rewarded for *sounding* right, not for *being* right. Which items in a history article — dates, names, places — are exactly the kind of thing it could get confidently wrong? Think about who or what could confirm each one, and at what point a person should step in.

🌎 In a New Context

A student uploads a photo of a handwritten note and asks an AI assistant to read it aloud and then summarize it in text. Name the property that lets one model move between image, text, and audio, and explain in one line how the network can handle all three.

❓ Lesson Question — Answered

The lesson asked how a language model generates text, what makes a model "large," and why it can still get facts wrong. A language model learns from large amounts of text to predict the likely next word, and it generates text by repeating that prediction, connecting one word at a time into a whole sentence. It chooses each word by a probability of coming next — the statistically plausible word, not the guaranteed-correct one — which is exactly why hallucination (fluent but false responses) can occur. A large language model scales this up: trained on vast Internet text with billions to trillions of parameters, and beyond a certain scale it gains qualitatively new abilities such as reasoning, summarizing, translating, and generating code (ChatGPT, Claude, and Gemini are examples). Because raw training does not guarantee helpful answers, RLHF tunes the model with human evaluations as rewards, making responses natural, on-intent, and safer. And multimodality lets a single model handle text, images, and audio together by converting them all into a common numerical representation. The lasting caution: fluency is not truth, so outputs must still be checked.

Exercise

1 Cover the Point! section with a red plastic sheet and quiz yourself in your notebook in numerical order.

2 Read the following passage and answer each question.

A model that predicts which word is likely to come next in a sentence is called a (a). Text generation is achieved by repeating the prediction of the (b). A language model based on a very large neural network trained using vast

⭐ REFLECT & CHALLENGE

Reflect: When you last used a generative-AI tool, did you check its facts? What could a confident wrong answer have cost? **Challenge:** A classmate says "RLHF makes the AI always tell the truth." Explain why that is not correct, and state what RLHF actually improves.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

amounts of text from the Internet is called a (c); when the number of (d) or the amount of training data exceeds a certain scale, qualitatively new abilities sometimes appear. Since just learning from vast text does not necessarily lead to responses desirable for humans, (e), a method of tuning by reinforcement learning using human evaluations as rewards, is used. Also, the property of being able to handle multiple kinds of data—text, images, audio, video, and so on—with a single model is called (f).

1. Fill in the blanks (a)–(f).

3 Answer the following questions.

(1) From the following options (A–D), choose the one that most appropriately describes the mechanism of text generation.

A Text is generated by connecting the results of repeatedly predicting the next word.

B Text is generated by selecting from thousands of prepared templates.

C Text is generated by first deciding the length of the sentence and then deciding all words simultaneously.

D Text is generated by machine-translating the input sentence.

(2) From the following options (A–D), choose the one that most appropriately describes the background to large language models acquiring new abilities.

A Because reducing the number of parameters below a certain level simplified the processing.

B Because limiting training data to a small amount of high-quality text raised quality.

C Because qualitatively new abilities appear when the number of parameters or the amount of training data exceeds a certain scale.

D Because humans wrote down the rules of language one by one and taught them to the model.

(3) From the following options (A–D), choose the one that most appropriately describes the purpose for which RLHF is performed.

A To increase the number of words a language model can handle.

B To make a language model's responses natural and aligned with the intent of the question.

C To enable a language model to handle data other than text.

D To reduce the number of parameters used in a language model's calculations.

⭐ Key takeaway

Remember:
A language model generates text by predicting the next word (plausible, not always true → hallucination). LLMs use vast data and billions to trillions of parameters; RLHF tunes them by human feedback; multimodality unifies text, image, and audio. Fluency ≠ truth — always verify.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

Editorial Team

Project Lead & Editorial Director

Masaki Iisaka

Editors & Writers

Biichi Nakajima
Hiromi Fujii

Reviewers

Prof. Mohamed Sayed Farag.

Prof. Eman Serag El-Din Bakr.

Dr. Abeer Hamed Ahmed.

Dr. Taher Abdel Hamid El-Adly

Dr. Manal Ziada Abdel Fadil.

The Central Administration for Curriculum Development

General Supervision

Dr. Gebriel Anwer Hemieda

Head of the Central Administration for Curriculum Development

All Rights Reserved ©Ministry of Education and Technical Education - 2026 / 2027

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means-electronic, photocopying, recording, or otherwise-without prior written permission of the Ministry of Education and Technical Education.

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

EDUCATION SPRINGS NEW LIFE

SPRIX

CACD

class	name

D.M.RAZK موقع الدكتور محمد رزق معلم الكيمياء التعليمي

موقع الدكتور محمد رزق التعليمي

امسح الباركود بالموبيل لتثبيت التطبيق وتحميل كل الملازم مجانا وبسرعه

MR

حمل التطبيق الأن

GET IT ON Google Play

ملازم ابتدائي | ملازم اعدادي | ملازم ثانـــوي | كورسات ونتائج امتحانات

موقع الدكتور محمد رزق
Dr.MoRizK